

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: June 17, 2003, 09:52:43 ; Search time 282 Seconds  
(Without alignments)  
2866.906 Million cell updates/sec

Title: US-09-972-758A-2  
Perfect score: 1910  
Sequence: 1 MAEPFLSEYOHQPOTSNCCTG.....LTENELHROERAPLSKFGD 359

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n model -DEV=xlh  
-O=/cgg2\_1/USPTO.spool/US09972758/r/nat\_17062003\_095237\_24364/app\_query.fasta\_1.519  
-DB=N\_Geneseq\_101002 -QPM=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosomb2 -TRANS=human40.cdi  
-LIST=45 -DOCALLIN=200 -THR\_SCORE=pc -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USR=US09972758.@CGM.1.1.263.erna\_17062003.095237.24364 -NCPU=6 -ICPU=3  
-NO\_MMP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_101002: \*  
1: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1980.DAT: \*  
2: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1981.DAT: \*  
3: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1983.DAT: \*  
4: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1984.DAT: \*  
5: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1985.DAT: \*  
6: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1986.DAT: \*  
7: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1987.DAT: \*  
8: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1988.DAT: \*  
9: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1989.DAT: \*  
10: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1990.DAT: \*  
11: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1991.DAT: \*  
12: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1992.DAT: \*  
13: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1993.DAT: \*  
14: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1994.DAT: \*  
15: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1995.DAT: \*  
16: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1996.DAT: \*  
17: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1997.DAT: \*  
18: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1998.DAT: \*  
19: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1999.DAT: \*  
20: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2000.DAT: \*  
21: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT: \*  
22: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2001B.DAT: \*  
23: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2001B.DAT: \*  
24: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2002.DAT: \*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1910	100.0	2199	19	AAV82778	Clone bp783_3 isol
2	1910	100.0	2195	24	ABO52015	Human polynucleoti
3	1903	99.6	2178	22	AAH18098	Human cDNA sequenc
4	1902	99.6	1080	24	ABN84013	Human oestrogen do
5	733	38.4	1083	24	ABL90067	Human polynucleoti
6	714	37.4	523	20	AAV86036	EST clone B227. H
7	629	32.9	461	24	ABL63543	Breast cancer rela
8	308.5	16.2	997	22	AAH96667	Human protein enco
9	245	12.8	342	20	AAV87316	EST clone Bp783.
10	214	11.2	1489	23	ABL06283	Drosophila melanog
11	205.5	10.8	3714	23	ABL06282	Drosophila melanog
12	184	9.6	549	21	AAV95217	Cat flea head and
13	174	9.1	1024	23	AAV94493	DNA encoding novel
14	162	8.5	7516	22	AAK51987	Human polynucleoti
15	159.5	8.4	4041	24	ABO54663	Human ovarian anti
16	159.5	8.4	7736	23	AAV65910	DNA encoding novel
17	158.5	8.3	9531	20	AAI22301	DNA encoding a hu
18	156	8.2	6432	23	AAV65903	Human telomerase p
19	155	8.1	2277	19	AAV13836	Homo sapiens mamma
20	155	8.1	2277	19	AAV05372	Human telomerase p
21	153.5	8.0	3579	23	ABO56551	Drosophila melanog
22	153.5	8.0	6105	23	ABO56550	Drosophila melanog
23	152	8.0	2421	24	ABR80600	Bacillus clausii g
24	152	8.0	3540	23	ABL06305	Drosophila melanog
25	151	7.9	3707	23	ABL06276	Drosophila melanog
26	149.5	7.8	8486	22	AAK52971	Human polynucleoti
27	149	7.8	2483	23	ABL07559	Drosophila melanog
28	149	7.8	4483	23	ABL07558	Drosophila melanog
29	149	7.8	18737	23	ABL07492	Drosophila melanog
30	149	7.8	18737	23	ABL07530	Drosophila melanog
31	148.5	7.8	7741	22	AAI60155	Human polynucleoti
32	148.5	7.8	7741	22	AAI60155	Human polynucleoti
33	148	7.7	5943	22	AAV22920	DNA encoding novel
34	147.5	7.7	3737	21	AAI18028	Lung cancer associ
35	147.5	7.7	15606	23	ABL16881	Drosophila melanog
36	147	7.7	7501	22	AAI58369	Human polynucleoti
37	147	7.7	7501	22	AAI58370	Human polynucleoti
38	145	7.6	4301	24	ABK63759	Rat sequence diffe
39	145	7.6	8731	23	ABL10296	Drosophila melanog
40	144	7.5	4714	22	AAV21393	Human cockayne syn
41	143.5	7.5	2159	15	AAV67223	Mouse p53MNC. Mus
42	143	7.5	867	23	AAV84549	DNA encoding novel
43	143	7.5	3399	22	AAV852051	Human polynucleoti
44	142	7.4	2818	23	AAV84552	DNA encoding novel
45	141	7.4	2304	12	AAO14176	Clone pXRc8 encod

ALIGNMENTS

RESULT 1  
AAV82778  
ID AAV82778 standard; cDNA: 2199 bp.  
XX AAV82778:  
AC  
AC  
DT 25-FEB-1999 (first entry)  
XX  
DE Clone bp783\_3 isolated from human foetal kidney cDNA library.  
XX  
XX Secreted protein: nutritional activity; immune stimulating; vaccine;  
XX suppressing activity; haematopoiesis regulating activity;  
XX tissue growth activity; activin; inhibin activity; chemotactaxis;  
XX chemokinetic activity; haemostasis; thrombolytic activity; receptor;  
XX ligand; anti-inflammatory; cadherin; tumour invasion suppressor;  
XX tumour inhibition; gene therapy; ds.  
XX  
XX Homo sapiens.  
OS

XX MO9842739-A2.  
PN  
XX  
PD 01-OCT-1998.  
XX  
PF 20-MAR-1998; 98WO-0505653.  
XX  
PR 19-MAR-1998; 98US-0044466.  
PR 21-MAR-1997; 97US-0822167.  
XX  
PA (GEMV ) GENETICS INST INC.  
XX  
PI Agostino MJ, Jacobs K, Lavallie BR, McCoy JM, Merberg D;  
PI Racie LA, Spaulding V, Treacy M;  
XX  
DR WPI: 1998-609890/51.  
XX  
DR P-PSDB; AAM85455.  
XX  
PT New polynucleotides encoding secreted human proteins - derived from  
PT human foetal brain, adult brain, foetal kidney, placenta or adult  
PT pineal gland cDNA libraries.  
XX  
PS Claim 1: Page 66-67; 113pp: English.

XX The present sequence encodes a secreted protein. The polynucleotide and  
CC secreted protein are predicted to have biological activities which would  
CC make them suitable for treating, preventing or ameliorating medical  
CC conditions in humans and animals, although no supporting data is given.  
CC Suggested activities include nutritional activity, immune stimulating  
CC (e.g. as vaccines) or suppressing activity, haematopoiesis regulating  
CC activity, tissue growth activity, actinin/inhibin activity,  
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,  
CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour  
CC invasion suppressor activity, and tumour inhibition activity (no data is  
CC given in the specification to support these activities). The  
XX polynucleotide is also stated to be useful for gene therapy.

SQ Sequence 2199 BP: 552 A; 511 C; 674 G; 462 T; 0 other;

Alignment Scores:

Pred. No.: 6.19e-106 Length: 2199  
Score: 1910.00 Matches: 359  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 19 Gaps: 0

US-09-972-758a-2 (1-359) x AAV82778 (1-2199)

QY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20  
DB 707 ATGCGGAGGACCTTGTGACAAATATCAACACGACCTCAAACTAGCACTGTACAGCT 766  
QY 21 AlaAlaAlaValGlnGlnGluLeuAsnProGlnArgProProGlyAlaGlnGluAlaGly 40  
DB 767 GCTGCTGCTCTCCAGGAAGAGCTGAACCTGAGCGGCCCCAGCGGGGAGCGGGT 826  
QY 41 ProGlnGluAspSerArgTTPGlnSerArgAlaPheProGlnLeuGlyAlaGProGly 60  
DB 827 CCGGAGAGGACAGTACGTGCGCAATGAGAGCGTTCCCGCAAGTTGGTGGCCCTCCGGG 886  
QY 61 ProGlnGlyGlnGlySerLeuGlnSerGlnProProProLeuGlnThrGlnAlaCysPro 80  
DB 887 CCGAGAGGGGAGAGGAGCGCTGGAATCCCAACCACTCCCTTGGACAGCCAGGCGCTGCCA 946  
QY 81 GluSerSerCysLeuArgGlnGlnGlyGlnGlyGlnAsnGlyGlnAspSerSerAlaGly 100  
DB 947 GAATCTAGCTGCTCTGAGAGAGGCGGAGAAAGGCGCAAAATGCGACACTGCTCGCTGGC 1006  
QY 101 GlyAspPheProProAlaGlnValGlnProThrProGlnAlaGlnLeuLeuAlaGln 120  
DB 1007 GCGGACTTCCCGCGCGCGGAGAGAGTGGAAACGACGCCCGAGGCTGCTGCCACG 1066

QY 121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlnGlnGlnGlu 140  
DB 1067 CTTGTGATACCTCCGACGCGCACTAAGTTGGGGCTCTCCCGCAGGGGGCGAAGAGAG 1126  
QY 141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgProSerLysLys 160  
DB 1127 TGGGACAGCAGCAGAGACAGCTGGCGAAGAAAACATAGAGAGCGCCGTCAGAGAG 1186  
QY 161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGlnGluLysLysLysPheAspGlu 180  
DB 1187 AAGCGCATTTGGAAACCGTACTGACAGCTGACCTGGGAGAGAGAAAAGTTGACAGAG 1246  
QY 181 LysGlnSerLeuArgAlaSerArgAlaAlaAlaLysSerAspSerAspThrSerAspAsp 200  
DB 1247 AAGACAGCGCTTGAGCTTCAAGATCCAGAGCCGACAGTGTCCCAAGGGCCAGCGCGTC 1306  
QY 201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluProAspLeu 220  
DB 1307 GCCCGCTATACACCCAGCGAGTTCCTCATGATGATCAGACACGAGAGGCGCGATCTC 1366  
QY 221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspSerAspThrSerAspAsp 240  
DB 1367 AAAACCGGCTGTACTCTCAAGCGCGCGCGCCCAAAATCCAGACACACGCGATGACAGC 1426  
QY 241 PheMetGlnGlnGlnGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260  
DB 1427 TTCATGGAAGAACAGGGGGTGGAGAGATGGGGCGCAGGATGGATGGAGGGAGCGGCAC 1486  
QY 261 GluPheLeuGlnArgAspPheSerGluThrTyrGlnArgTyrHisThrGlnSerLeuGln 280  
DB 1487 GAATTTCTGACGGGGAGCTTCTGGAGACCTAGACCGGATACACACGGAGACCTGCGAG 1546  
QY 281 AsnMetSerLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300  
DB 1547 AACATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1606  
QY 301 MetGluAspGlnAsnAsnArgLeuArgLeuGlnSerLysArgLeuGlnGlyLysAspAspAla 320  
DB 1607 ATGAGAGACGACGAACAACCGCGCTGGAGAGCAAGCGCGTGGCGCGACGACGCGC 1666  
QY 321 ArgValArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 340  
DB 1667 CGTGTGCGGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGG 1726  
QY 341 ThrGluAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 359  
DB 1727 ACCGAGACGAACTGACCGCGCGACGAGCGAGCGCGCGCTTCCAAAGTTTGGAGAC 1783

RESULT 2  
ABQ92015 standard: cDNA: 2199 BP.  
ID ABQ92015.  
XX  
AC ABQ92015.  
XX  
DT 04-OCT-2002 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 12.  
XX  
KW Human: cytostatic; antipneumatic; antiarthritic; vulnery; analgesic;  
KW antiinflammatory; antibacterial; immunosuppressive; antiparkinsonian;  
KW neuroprotective; nootropic; osteopathic; haemostatic; vasotropic;  
KW antilicer; fungicide; antidiabetic; antisthmatic; antiallergic;  
KW immunostimulant; antiparasitic; secreted protein; transmembrane protein;  
KW cytokine; cell proliferation; cell differentiation; autoimmune disease;  
KW stem cell; growth factor; nervous system disease; neuropathy;  
KW Alzheimer's disease; Parkinson's disease; Huntington's disease;  
KW osteoporosis; severe combined immunodeficiency; SCID; infection;  
KW multiple sclerosis; rheumatoid arthritis; gene therapy; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN US2002065394-A1.  
XX







ABN84013  
ID ABN84013 standard; cDNA; 1080 BP.  
XX  
AC ABN84013:  
XX  
DT 10-SEP-2002 (first entry)  
XX  
DE Human oestrogen downregulated gene EDG1 coding sequence.  
XX  
XX EDG1: oestrogen downregulated gene; tumour suppressor; human:  
KM breast cancer; prostate cancer; testicular cancer; ovarian cancer;  
KM uterine cancer; colon cancer; chromosome 17q; gene therapy; gene:  
ss.  
XX Homo sapiens.  
XX  
FH Key Location/Qualifiers  
XX CDS 1..1080  
FT /\*tag= a  
FT /product= "EDG1"  
FT /transl\_except= "(pos:460..462,aa:Arg)"  
FT /transl\_except= "(pos:511..513,aa:Thr)"  
XX  
XX MO200228879-A1.  
XX  
XX 11-APR-2002.  
XX  
XX 05-OCT-2001: 2001MO-US31300.  
XX  
XX 05-OCT-2000: 2000US-238187P.  
XX  
XX (UYCA-) UNIV CASE WESTERN RESERVE.  
XX  
XX Montano M, Wiltman B:  
PI WPI: 2002-519107/55.  
DR P-PSDB: ABB76495.  
XX  
XX Polynucleotides encoding Estrogen Down-Regulated Gene 1 proteins,  
PT useful for the prevention, diagnosis and treatment of e.g. breast  
PT cancer, testicular cancer, prostate cancer, uterine cancer, cervical  
PT cancer, ovarian cancer and colon cancer -  
XX  
XX Claim 1: Fig 1A-B; 52pp; English.  
XX  
XX The present sequence is the coding sequence for human oestrogen  
CC downregulated gene 1 (EDG1), a tumour suppressor gene that is  
CC downregulated by oestrogen in mammary epithelial cells. The gene  
CC was identified by yeast two-hybrid screenings for oestrogen  
CC receptor-interacting proteins in breast epithelial cells. It  
CC was localised to chromosome arm 17q. EDG1 mRNA expression is  
CC prevalent in normal mammary epithelial cells and in other human  
CC hormone-responsive tissues such as the ovary, prostate and testis.  
CC Expression is low in breast cancer epithelial cells. Oestradiol,  
CC which induces breast cancer cell growth, has an inhibitory effect  
CC on EDG1 mRNA expression in breast cancer cells. Hexamethylene  
CC bis-acetamide, an inducer of differentiation and apoptosis,  
CC upregulates EDG1 mRNA expression in breast cancer cells. The  
CC invention provides EDG1 polynucleotides and polypeptides. In a  
CC claimed method, a test sample from an individual suspected of  
CC having, or known to have breast, testicular, prostate, uterine,  
CC cervical, ovarian or colon cancer is assayed for EDG1 transcript  
CC using a polynucleotide that is complementary to the present  
CC sequence or by RT-PCR using a primer derived from the present  
CC sequence. A decrease in the level of transcript compared to the  
CC level in a test sample indicates that the test sample contains or  
CC was derived from cancerous cells antibody. A claimed method for  
CC decreasing the proliferation of breast, prostate, testicular,  
CC ovarian, uterine, cervical or colon cancer cells involves increasing  
CC EDG1 protein activity in the cells, either by contacting the cells  
CC with EDG1 protein or its fragment or functional equivalent, or with  
CC a nucleic acid encoding EDG1 protein, its fragment or functional  
CC equivalent.

XX  
SQ Sequence 1080 BP; 265 A; 296 C; 376 G; 143 T; 0 other:  
Alignment Scores:  
Pred. No.: 9.01e-106 Length: 1080  
Score: 1902.00 Matches: 357  
Percent Similarity: 99.72% Conservative: 1  
Best Local Similarity: 99.44% Mismatches: 1  
Query Match: 99.58% Indels: 0  
DB: 24 Gaps: 0  
US-09-972-758A-2 (1-359) x ABN84013 (1-1080)  
QY 1 MetAlaGluProPheLeuSerGluTrpGlnHisGlnProGlnThrSerAsnGlyThrGly 20  
|||||  
Db 1 ATGGCCGAGCGCAATCTTGTGACAGATATCAACACCACCTCAAACTGACAACTGTACAGGT 60  
21 AlaAlaAlaValGlnGlnGlnLeuAsnProGluArgProProGlyAlaGlnGlnArgVal 40  
|||||  
Db 61 GCTGCTGCTGTCTCCAGGAAGAGTGAACCTTGACGGCCCCCAGCGCGGAGAGCGGGTG 120  
41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60  
|||||  
Db 121 CCCGAGGAGGACAGTAgGTGGCAATCCAGAGCGTTCGCCAGTGGGTGGCGTCGGGG 180  
61 ProGluGluGlySerLeuGlnSerGlnProProProLeuGlnThrGlnAlaGlyPro 80  
|||||  
Db 181 CCGGAGGGGAGGAGGAGCGCTCGAATCCCAACCACTCCCTTGAGAGCCGAGGCTGTCCA 240  
81 GluSerSerCysLeuArgGlnGlnGlyGlyGlyGlnAsnGlyAspAspSerSerAlaGly 100  
|||||  
Db 241 GAATCTAGCTGCTGTGAGAGAGGCGGAGAGGGCCAGAAATGGGAGCAGCTGCTCGCTGGC 300  
101 GlyAspPheProProProAlaGlnValGlnProThrProGluAlaGlnLeuLeuAlaGln 120  
|||||  
Db 301 GCGGACCTTCCCGCGCGCGGAGAGTGAACCCAGCCGCGAGCGCTGCTCGCCAG 360  
121 ProCysHisAspSerGlnAlaSerLysLeuGlyAlaProAlaAlaGlnGlyGlnGln 140  
|||||  
Db 361 CCTTGTCATGACTCCAGAGCGCACTAAGTTGGGGCTCTGCGCCAGAGGGCGAAGAGGAG 420  
141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgProSerLysLys 160  
|||||  
Db 421 TCGGGACAGACGACGACAGCTGCGGAGAAACAAATAGAGAGCCCTTCACAGAG 480  
161 LysArgHisTrpLysProTrpTrpLysLeuThrTrpGlnGlnLysLysPheAspGln 180  
|||||  
Db 481 AACCGGCATTGGAAACCGTACTCAAGCTGAAGTGGGAAGAGAAAGATTTCAGAGAG 540  
181 LysGlnSerLeuArgAlaSerArgIleArgAlaGlnMetPheAlaLysGlnGlnProVal 200  
541 AAACAGAGCGCTTCGACTTCAGATCCGAGCGCAATGTTCGCCCAAGGGCCGCGGTC 600  
201 AlaProTrpAsnThrThrGlnPheLeuMetAspHisAspGlnGlnGlnProAspLeu 220  
601 GCCCCTTAATACACCCAGCGAGTTCCTCATGGATGATCAGACGAGGAGCGGATCTC 660  
221 LysThrGlnLeuTrpSerLysArgAlaAlaLysSerAspAspThrSerAspAsp 240  
|||||  
Db 661 AAACCGGCGCTGTACTCAAGCGGGCGCGCCCAATTCGAGACACACCGCATGACAGC 720  
241 PheMetGlnGlnGlnGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260  
721 TTCATGAGGAAGAGGGGTGAGGAGATGGGGCGAGCATGGATGGAGGGAGCGGCAC 780  
261 GlnPheLeuGlnArgAspPheSerGlnTrpTrpGluArgTrpHisThrGlnSerLeuGln 280  
|||||  
Db 781 GAGTTTCTCAGCGGAGCTTCTCGAGAGCTAGAGACGGGTACACACGAGAGACCTCGCAG 840  
281 AsnMetSerLysGlnGlnLeuLeuLysGlnTrpTrpGlnGlnGlnGlnGlnGlnGlnGln 300  
|||||  
Db 841 AACATGAGCAAGGAGGAGCTATCAAGAGTACTGTGAAGTGGAGAGAGTCTTCGCGC 900

QY 301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyIleAspAspAla 320  
 Db 901 ATGGAGGAGCAAGAACCGCGCTGGCTGAGAGCAAGCGGCTGGCTGGAGACGCG 960  
 QY 321 ArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340  
 Db 961 CGTGTGGGAGAGCTGGAGCTGAGCTGACCGCGCTGGCGCGGAGCACTCCAGCTCTG 1020  
 QY 341 ThrGluAsnGluLeuHisArgGlnGlnGlnArgAlaProLeuSerLysPheGlyAsp 359  
 Db 1021 ACCGAGACGACACTGCAACCGGACAGAGAGACCGCGCTTCCAGTTTGAGAC 1077  
 RESULT 5  
 ABL90067/c  
 ID ABL90067 standard; cDNA; 1083 BP.  
 AC ABL90067;  
 DT 24-MAY-2002 (first entry)  
 XX Human polynucleotide SEQ ID NO 629.  
 DE  
 XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
 KW vulnerrary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein; gene; ss.  
 XX Homo sapiens.  
 OS  
 XX NC0200190304-A2.  
 PN 29-NOV-2001.  
 PD  
 XX 18-MAY-2001; 2001MO-US16450.  
 PF  
 XX 19-MAY-2000; 2000US-205515P.  
 PR  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA Birse CE, Rosen CA;  
 PI  
 XX WPI: 2002-122018/16.  
 DR P-PSDB; ABB89658.  
 XX  
 XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
 PT prevention of neural, immune system, muscular, reproductive,  
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
 PT disorders -  
 PS  
 PS Claim 4: SEQ ID NO 629; 2081pp + Sequence Listing; English.  
 XX  
 XX The invention relates to novel genes (ABL89449-ABL90853) and proteins  
 CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO. int/pub/publicated\_pct\_sequences.  
 XX  
 XX Sequence 1083 BP; 301 A; 290 C; 199 G; 286 T; 7 other;

Alignment Scores:  
 Pred. No.: 1.16e-35 Length: 1083  
 Score: 733.00 Matches: 146  
 Percent Similarity: 98.65% Conservative: 0  
 Best Local Similarity: 98.65% Mismatch: 2  
 Query Match: 38.38% Indels: 1  
 DB: 24 Gaps: 0  
 US-09-972-758A-2 (1-359) x ABL90067 (1-1083)  
 QY 212 AsHisAspGlnGluGluProAspLeuLysThrGlyLeuTySerLysArgAlaAlaAla 231  
 Db 1081 GATCAGCAGCAGAGAGAGCGGACCTCAAAACCGGCTTACATCCARCGGSCCGCGC - 1023  
 QY 232 LysSerAspAspThrSerAspAspAspPheMetGluGluGlyGluGluAspGlyGly 251  
 Db 1022 AAATCCGACGACACACGACGATGACGACTTATGAGAGAGGGGTGAGAGATGGCGCG 963  
 QY 252 SerAspGlyMetGlyLysAspGlySerGluPheLeuGlnArgAspPheSerGluThrTy 271  
 Db 962 AGCGATGGATGGAGAGGAGCGACGCGACGAGCTTCTCGAGCGGGAGCTTCTCGAGACGCTAC 903  
 QY 272 GluArgThrIshThrGluSerLeuGlnAsnMetSerLysGlnGluLeuIleLysGluTy 291  
 Db 902 GAGCGGTACCAACAGAGAGACCTTCAGACATGAGCAAGAGAGAGCTCATCAAGAGACTAC 843  
 QY 292 LeuGluLeuGluLysCysLeuSerArgMetGluAspGluAsnAsnArgLeuArgLeuGlu 311  
 Db 842 CTGGAACTGAGAAAGTCCCTCTCGCGCATGAGAGACAGAAACCGGCTGGCGGTGGAG 783  
 QY 312 SerLysArgLeuGlyGlyAspAspAlaArgValArgGluLeuGluLeuAspArg 331  
 Db 782 AGCAAGCGCGCTGGTGGAGAGACCCCGCTGCGGAGCGTGGAGCGTGGAGCGG 723  
 QY 332 LeuArgAlaGluAsnLeuGlnLeuLeuThrGluAsnGluLeuHisArgGlnGlnArg 351  
 Db 722 CTGGCGCGCGAGAGACCTTCAGCTGCTGACCGAGAGACCACTGACCGGAGAGAGCGA 663  
 QY 352 AlaProLeuSerLysPheGlyAsp 359  
 Db 662 GCGCGGCTTCCAAAGTTTGAGAC 639  
 RESULT 6  
 AAV86036  
 ID AAV86036 standard; cDNA; 523 BP.  
 XX  
 XX AAV86036;  
 AC  
 XX 27-APR-1999 (first entry)  
 DT  
 XX  
 XX EST clone B227.  
 DE  
 XX Expressed sequence tag; secreted protein; haematopoiesis regulator;  
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;  
 KW chemokinesis; chemokinesis; haemostasis; gene therapy; thrombolytic;  
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.  
 OS  
 XX Homo sapiens.  
 XX  
 XX NC09845435-A2.  
 PN 15-OCT-1998.  
 PD  
 XX 10-APR-1998; 98MO-US06954.  
 PF  
 XX 10-APR-1997; 97US-0835913.  
 PR  
 XX (GEMV ) GENETICS INST INC.  
 PA  
 XX Agostino M.J. Jacobs K. Lavallie ER, McCoy JM, Merberg D;  
 PI Racie LA, Spaulding V, Treacy M;  
 XX  
 XX WPI: 1999-070076/06.

```
XX New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries
XX
XX Claim 1: Page 97: 633pp; English.
XX
CC This sequence represents an expressed sequence tag (EST), and is a
CC polynucleotide of the invention. The polynucleotides of the invention are
CC all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemokine/chemokine activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene
CC therapy.
XX
SQ Sequence 523 BP; 112 A; 149 C; 153 G; 108 T; 1 other:
XX
Alignment Scores:
Pred. No.: 7-61e-35 Length: 523
Score: 714.00 Matches: 131
Percent Similarity: 99.268 Conservative: 3
Best Local Similarity: 97.048 Mismatches: 1
Query Match: 37.384 Indels: 0
DB: 20 Gaps: 0
US-09-972-758a-2 (1-359) x AAV86036 (1-523)
OY 1 MetaAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
DB 104 ATGGCCGAGCCATTCTTGTTCAGATATCAACACCCAGCTCAAACTGACAGTGTACAGCT 163
OY 21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProGluValGluGluVal 40
DB 164 GCTGCTGCTGTCCAGAAAGCTGAACCTGAGCGCCGCCGAGGAGGAGGAGG 223
OY 41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyArgProGly 60
DB 224 CCCGAGGAGGACAGTAGTGTCGCAATGACAGCGCTCCCGAGTGGGTGGCCCTCGGG 283
OY 61 ProGluGluGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
DB 284 CCGGAGCGGGAAGGAGCGCTGGAATCCCAACCACTCCCTTGAGACCCAGCGCTGCCA 343
OY 81 GluSerSerCysLeuArgGluGlyGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 100
DB 344 GAATCTAGCTGCTGTGAGAGAGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 403
OY 101 GlysPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
DB 404 GCGGACTTCCGCCCGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 463
OY 121 ProGlyHisAspSerGluAlaSerLysLeuGlyAlaProAlaAla 135
DB 464 CCTTGTCAATGACTCCGAGCGCAAGTAAGATGGGTGACGCGCTCT 508
RESULT 7
ABL63543
ID ABL63543 standard; DNA: 461 BP.
XX
XX ABL63543:
XX
XX 15-MAY-2002 (first entry)
XX
XX Breast cancer related gene sequence SEQ ID NO:1880.
XX
XX Human: cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
```

```
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
XX Homo sapiens.
XX
XX MO200194629-A2.
XX
XX 13-DEC-2001.
XX
XX 30-MAY-2001: 2001WO-US10838.
XX
XX 05-JUN-2000: 2000US-209473P.
XX 05-JUN-2000: 2000US-209531P.
XX 18-SEP-2000: 2000US-23313P.
XX 18-SEP-2000: 2000US-233617P.
XX 20-SEP-2000: 2000US-234009P.
XX 20-SEP-2000: 2000US-234034P.
XX 20-SEP-2000: 2000US-234052P.
XX 22-SEP-2000: 2000US-234509P.
XX 22-SEP-2000: 2000US-234567P.
XX 25-SEP-2000: 2000US-234923P.
XX 25-SEP-2000: 2000US-234924P.
XX 25-SEP-2000: 2000US-235077P.
XX 25-SEP-2000: 2000US-235082P.
XX 25-SEP-2000: 2000US-235134P.
XX 25-SEP-2000: 2000US-235280P.
XX 26-SEP-2000: 2000US-235637P.
XX 26-SEP-2000: 2000US-235638P.
XX 27-SEP-2000: 2000US-235711P.
XX 27-SEP-2000: 2000US-235720P.
XX 27-SEP-2000: 2000US-235840P.
XX 27-SEP-2000: 2000US-235863P.
XX 28-SEP-2000: 2000US-236028P.
XX 28-SEP-2000: 2000US-236032P.
XX 28-SEP-2000: 2000US-236033P.
XX 28-SEP-2000: 2000US-236034P.
XX 28-SEP-2000: 2000US-236109P.
XX 28-SEP-2000: 2000US-236111P.
XX 29-SEP-2000: 2000US-236842P.
XX 29-SEP-2000: 2000US-236891P.
XX 02-OCT-2000: 2000US-237172P.
XX 02-OCT-2000: 2000US-237173P.
XX 02-OCT-2000: 2000US-237278P.
XX 02-OCT-2000: 2000US-237284P.
XX 02-OCT-2000: 2000US-237295P.
XX 02-OCT-2000: 2000US-237316P.
XX 03-OCT-2000: 2000US-237425P.
XX 03-OCT-2000: 2000US-237598P.
XX 03-OCT-2000: 2000US-237604P.
XX 03-OCT-2000: 2000US-237606P.
XX 03-OCT-2000: 2000US-237608P.
XX 01-NOV-2000: 2000US-244867P.
XX 01-NOV-2000: 2000US-245084P.
XX
XX (AVAL-) AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Edner R, Endress G, Horrigan S;
XX Sopfel DR, Weaver Z;
XX
XX MPI: 2002-188264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a
XX chemical agent to be tested for anti-neoplastic activity, and
XX determining a change in expression of a gene of a signature gene set -
XX
XX Claim 1: SEQ ID 1880: 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an
XX anti-neoplastic agent. The method involves exposing cells to a chemical
XX agent to be tested for anti-neoplastic activity, determining a change in
XX expression of at least one gene (I) of a signature gene set, where (I)
XX comprises a sequence (S) selected from 8447 sequences (given in ABL61664
```



Db 878 -----||| ||| ||||| |||  
GAGCGCGGACAGTGTGGCGCGGACCGAC 843  
Qy 260 SerGluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThGlnSerLeu 279  
Db 842 GGTAGATTCCAGCGGAAGACTCTGTGAGACTTACGACCTTCCACACCGAAGCCTG 783  
Qy 280 GlnAsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluGluGluLysCysLeuSer 299  
Db 782 CAGGCGCGCAGACAGACAGACTGTGTCGACTACCTGGAGCTGGAGAACCGGCTGTG 723  
Qy 300 ArgMetGluAspGluAsnAsnArgLeuArgLeuSerLysArgLeuGlyGlyAspAsp 319  
Db 722 CAGCGGAGGAGAGACTAGACGCGTGCAGCAGCTGCAGCGCGTGCAGCGGACGAC 663  
Qy 320 AlaArg---ValArgGluLeuGluGluGluLeuAspArgLeuArgAlaGluAsnLeuGln 338  
Db 662 TGCCGCGAGGTGAGAGACTGTGTCGCGAGCTCAGAGGCTCCGACCGAAACCGACGCG 603  
Qy 339 LeuLeuThrGluAsnGluLeuHisArgGlnGln 349  
Db 602 CTGCTGACGAGAACCAATGTGACACCGAGAG 570  
RESULT 9  
AAV87316  
ID AAV87316 standard; cDNA: 342 BP.  
XX  
AC AAV87316:  
XX  
DT 27-APR-1999 (first entry)  
XX  
DE EST clone BP783.  
XX  
KW Expressed sequence tag; secreted protein; haematopoiesis regulator;  
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;  
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;  
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.  
XX  
OS Homo sapiens.  
XX  
PN MO9845435-A2.  
XX  
PD 15-OCT-1998.  
XX  
PE 10-APR-1998; 98WO-US06954.  
XX  
PR 10-APR-1997; 97US-0835913.  
XX  
PA (GEMV ) GENETICS INST INC.  
XX  
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;  
PI Racie LA, Spaulding V, Treacy M;  
XX  
DR WPI: 1999-070076/06.  
XX  
PT New polynucleotides encoding human secreted proteins - derived from  
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
PT ovary, pituitary, retina and colon cDNA libraries  
XX  
PS Claim 1: Page 526; 633pp; English.  
XX  
CC This sequence represents an expressed sequence tag (EST), and is a  
CC polynucleotide of the invention. The polynucleotides of the invention are  
CC all secreted EST sequences isolated from a variety of human tissue  
CC sources. The EST sequences and proteins encoded by them are predicted to  
CC have useful biological activities which would make them suitable for  
CC treating, preventing or ameliorating medical conditions in humans and  
CC animals, although no supporting data is given. Suggested activities  
CC include nutritional activity, immune stimulating or suppressing activity,  
CC haematopoiesis regulating activity, tissue growth activity, haemostatic  
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition

CC activity. The EST sequences are also stated to be useful for gene  
CC therapy.  
XX  
SQ Sequence 342 BP; 90 A; 79 C; 83 G; 89 T; 1 other;  
Alignment Scores:  
Pred. No.: 6,61e-07 Length: 342  
Score: 245.00 Matches: 46  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 12.83% Indels: 0  
DB: Gaps: 0  
US-09-972-758A-2 (1-359). x AAV87316 (1-342)  
Qy 1 MetaLactiPPropHeLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThGly 20  
Db 196 ATGGCCGAGCATTTCTGTGCAAAATATCAACACCGAGCTCAAACTAGCACTGTACAGGT 255  
Qy 21 AlaAlaValAlaGlnGluGluLeuAsnProGluArgPropGluAlaGluArgVal 40  
Db 256 GCTGCTGCTGTCCAGAGAGACTTAACCTGTAGCGCCGCCCGAGCGCGGAGACGGGTG 315  
Qy 41 ProGluGluAspSerArg 46  
Db 316 CCCGAGGAGACAGTAGG 333  
RESULT 10  
ABL06283  
ID ABL06283 standard; cDNA: 1489 BP.  
XX  
AC ABL06283:  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13331.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PE 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI: 2001-656860/75.  
DR P-PSDB: ABB62180.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Claim 1: SEQ ID NO 13331; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL06176-ABL30511), expressed DNA  
CC sequences (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 1489 BP; 434 A; 363 C; 383 G; 309 T; 0 other;

# Alignment Scores:

Pred. No.:	Length:	Matches:	Mismatches:	Indels:	Gaps:
Score: 0.000216	1489	73	46	69	10
Percent Similarity: 214.00					
Best Local Similarity: 42.20%					
Query Match: 25.89%					
DB: 11.20%					

US-09-972-758a-2 (1-359) x ABL06283 (1-1489)

```

OY 85 LeuArgGluGlyGluLysGlyGlnAsnGlyAsp-----AspSerSerAla 99
DB 98 ATGGCTGAAGCTGTAAATAATGAAGTGGCTCCCAACAGACCTTTGGATAGCGGCGGA 157
OY 100 GlyGlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuAla 119
DB 158 GGAGGGGCC----- 166
OY 120 GlnProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaGlyGlyGlu 139
DB 167 -----GCCAGTGGCGGAGCGGCTGCTGTAAGAGCGGCTCG 205
OY 140 GluTrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLys 159
DB 206 GGAATGCCGAGAGGAGAAACATCGCGGGGAAAAAATCC---AAATGCGAGCCCAAGAG 262
OY 160 LysLysAlaGHisTrpLysProTyrTrpLysLeuThrTrpGlu-----GluLys 175
DB 263 ACCAAGAACCAATTAT---CCGACGTGAACCTGATATGTCACCGGTGACAGACCAACG 319
OY 176 LysLysPheAspGluLysGlnSerLeuArgAlaSerArgGlyLeuArgAlaGluMetPheAla 195
DB 320 CTGGAGGGGATCAACAGCAGACATAGCAGAACCAAGCTGGTTCGTGGGCTCTTG--- 376
OY 196 LysGlyGlnProValAlaProTyrAsnThrThrGlnPheLeuMetAspHisAspGln 215
DB 377 -----CTCGTTCGACACAGAACCGCTTCCTAATGAGAACACACTGTCG 424
OY 216 GluGluProAspLeuLysThrGlyLeuTyrSerLysAlaAlaAlaLysSerAsp 235
DB 425 GAG-----CTGCACAAACATGAC 442
OY 236 ThrSerAspAspPheMetGluGluGlyGlyGlnGluAspGlyGlySerAspGlyMet 255
DB 443 TCCGACGCAACTGTTT-----GGCTCCAGACCGAA----- 475
OY 256 GlyLysAspGlySerGluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHis 275
DB 476 -----GATCAGGTGCTCTTCTTCACAGAGATTCTCGATGCTCAGACGAGCGCG 529
OY 276 ThrGluSerLeuGlnAsnMetSerLysGlnGluLeuLeuLysGluTyrLeuGluGln 295
DB 530 CTCGACGCTTGTGACGATGATGCGCAAGCAGAGCTCATCCAGAGTGCATCAATGCA 589
OY 296 LysCysLeuSerArgMetGluAspGluAsnAsnArgLeuArgLeuGlySerLysArgLeu 315
DB 590 GATGATATTTCGAGGCCCAACATATCCAGAGATT-----GGAGCCAAATTA 640
OY 316 GlyLysAspAspAlaArgValArgGluLeuGluLeuLeuAspArgLeuArgAlaGlu 335
DB 641 CGACGCGAGATGATAGATTCGCCCACTATCAGCTGAGATTCATTCTCGCAGCCAC 700
OY 336 AsnLeuGlnLeuLeuThrGluAsnGlu-LeuHisArgGlnGlnGluArgAlaProLeuSe 355
DB 701 CTCCTTGCGCACGTTTCCGCGACAGCTGACCAAGCGCAGCAGCGGCGCGACTTCC 760
OY 355 rLys 356
  
```

DB 761 CCGC 764

## RESULT 11

ABL06282/C  
 ID ABL06282 standard; cDNA; 3714 BP.

AC ABL06282;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13328.

KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI: 2001-656860/75.

P-PSDB: ABB62179.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions.

PS Claim 1; SEQ ID NO 13328; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (AB57737-AB872072).

CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 3714 BP; 971 A; 891 C; 852 G; 1000 T; 0 other;

## Alignment Scores:

Pred. No.:	Length:	Matches:	Mismatches:	Indels:	Gaps:
Score: 0.00178	3714	69	42	79	10
Percent Similarity: 205.50					
Best Local Similarity: 40.36%					
Query Match: 25.09%					
DB: 10.76%					

US-09-972-758a-2 (1-359) x ABL06282 (1-3714)

```

OY 79 CysProGluSerSerCysLeuArgGluGlyLysGlyGlnAsnGlyAsp----- 95
DB 2507 TGCAGTGGGCGAGAGTTGATTTTGGAGTATTTTGGCGGTTCAAAFTTGAATTGGCTCTC 2448
OY 96 -----AspSerSerAlaGlyGlyAsp 102
DB 2447 TTGCTCTTTCAGAAAGTGGCTCCCAACAGACCTTTGGATTACGGCGGAGGAGGGGC 2388
OY 103 PheProProProAlaGluValGluProThrProGluAlaGluLeuAlaGlnProCys 122
  
```



```

Db      2388 ----- 2388
QY      123 HisaspsergluaIsertLysleuGlyAlaProAlaIaGlyGluGluGluTrpGly 142
Db      2387 -----GCCAGTGGCGGAGGCGGTCTCTAGAGGCGGCTCGGAATCCG 2340
QY      143 GlnGlnGlnArgGlnLeuGlyLysLysHisArgArgProSerLysLysArg 162
Db      2339 AAGGGAACATCGCGCGGGAATAATCC---AAGATGACGCCCAAGAACACCAAGAAC 2283
QY      163 HisTrpLysProTyrTyrLysLeuThrTrpGlu-----GluLysLysPhe 178
Db      2282 CATATAT---CCGCGATGGAAATCGATATGTCACAGCGGTGACAGGACCGCTGAGGCG 2226
QY      179 AspGluLysGlnSerLeuArgAlaSerArgLeuArgIaGluMetPheAlaLysGlyGln 198
Db      2225 AATCAACGACAGATATACAGAACCAACCTGCTTCCTCCGCCCTCTTG----- 2178
QY      199 ProValAlaProTyrAsnThrThrGlnPheLeuMetAspHisAspGlnGluPro 218
Db      2177 ---CTGCTTCGCTACACAGCAACCGCTCTCTATGAGACACATGCTCGAG----- 2127
QY      219 AspleuLysThrGlyLeuTyrSerLysArgAlaAlaLysSerAspAspThrSerAsp 238
Db      2126 -----CTGCACAAGATGACTCCGACGAC 2103
QY      239 AspspPheMetGluGluGlyGlyGluGluAspGlyLysSerAspGlyMetGlyLysP 258
Db      2102 AACGCTTTT-----GGCTCCAGACCGAA-----GAT 2076
QY      259 GlySerGlnPheLeuGlnArgAspPheSerGlnThrTyrGluArgTyrHisThrGluSer 278
Db      2075 CAGGTGCTCTCTCTCCAGAGAGTCTCCGATGTCTACAGAGCGGCGCTCGAAGCT 2016
QY      279 LeuGlnAsnMetSerLysGlnGluLeuLysLeuTyrLeuGluLeuLysCysLeu 298
Db      2015 TTGGAGACGATGACCAAGACGAGAGCTCATCCAGAGTGCATAAATCGAAGATCATAT 1956
QY      299 SerArgMetGluAspGluAsnAsnArgLeuArgLeuSerLysArgLeuGlyLysP 318
Db      1955 TCGAAGGCCAGAACATATCCACAGAGTTT-----GGAGCCAATATACGAGCGCAG 1905
QY      319 AspaLArgValArgGluLeuGluLeuGluLeuAspArgLeuArg 333
Db      1904 GATGATAGAGATGCCCAACTATACGTGAGATCATGATATGCCA 1860

RESULT 12
AAC95217
ID      AAC95217 standard; cDNA; 549 BP.
XX
AC      AAC95217:
XX
DT      19-FEB-2001 (first entry)
XX
DE      Cat flea head and nerve cord (HNC) cDNA, SEQ ID NO:1712.
XX
KW      Cat flea; head and nerve cord nucleic acid; HNC; flea infestation;
KM      vaccine; antiparasitic; therapeutic target; diagnosis; detection; ss.
XX
OS      Ctenocephalides felis.
XX
PN      MO200061621-A2.
XX
PD      19-OCT-2000.
XX
PF      07-APR-2000; 2000WO-US09437.
XX
PR      09-APR-1999; 99US-0128704.
XX
PA      (HESK-) HESKA CORP.
XX
PI      Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N:
XX

```

```

DR      WPI: 2000-656323/63.
XX
PT      Flea Malpighian tubule and head and nerve cord tissue derived nucleic
PT      acids useful for the prevention, diagnosis and treatment of flea
PT      infestations -
XX
PS      Claim 26; Page 818; 964pp; English.
XX
CC      The invention relates to novel cat flea (Ctenocephalides felis) nucleic
CC      acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
CC      or head and nerve cord (HNC) tissue. The invention also relates to the
CC      encoded proteins. The invention additionally encompasses expression
CC      constructs, recombinant viruses and recombinant cells comprising the
CC      nucleic acids of the invention, recombinant production of the proteins,
CC      antibodies against the proteins, a method of identifying inhibitors of
CC      the proteins, and compositions comprising the inhibitors for
CC      administration to an animal. The nucleic acids, and the proteins they
CC      encode may be used in the prevention, treatment and diagnosis of diseases
CC      associated with flea infestations. For example, the nucleic acids may be
CC      used to produce an HMT or HNC protein according to standard recombinant
CC      DNA methodology by inserting the nucleic acids into a host cell and
CC      culturing the cell to express the protein. The HMT and HNC nucleic acids
CC      may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
CC      and quantitate the presence of cat flea or other homologous nucleic acid
CC      sequences in samples. They may also be used to study the expression and
CC      function of the proteins and their role in metabolism. The HMT and HNC
CC      proteins may be used as antigens in the production of specific
CC      antibodies, and in assays to identify modulators (agonists and
CC      antagonists) of HMT and/or HNC protein expression and activity. The
CC      anti-HMT/HNC protein antibodies and antagonists may also be used to
CC      downregulate protein expression and activity. The antibodies may also be
CC      used as diagnostic agents for detecting the presence of flea polypeptides
CC      in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
CC      present sequence represents a cat flea HNC cDNA of the invention.
XX
SQ      Sequence 549 BP; 207 A; 100 C; 113 G; 126 T; 3 other:
XX

Alignment Scores:
Pred. No.: 0.00489 length: 549
Score: 184.00 Matches: 66
Percent Similarity: 42.29% Conservative: 30
Best Local Similarity: 29.07% Mismatches: 53
Query Match: 9.63% Indels: 78
DB: 21 Gaps: 12

US-09-972-758A-2 (1-359) x AAC95217 (1-549)
QY      34 ProGlyAlaGluGluArgValProGluGluAspSerArgTrpGlnSerArgAlaPhe--- 52
Db      33 CCAGCNGCTGAGAGAAATAGTG-----AGTAATATCATCATCAAGTCTTTT 80
QY      53 ---ProGlnLeuGlyArgArgProGlyProGluGlyGlnGlySerLeuGlnPro 71
Db      81 TTAACTACTGTTGAACATCAAAATGACGTGTAAATTTCA---AATATTGAATCGGAAT 137
QY      72 Pro-----ProLeuGlnThrGlnAla 78
Db      138 CCAAAATATATACACCGCAGTACAGCGTCCAAATCTGCCCCCTTCAACACAGCGCT 197
QY      79 CysProGluSerSerCysLeuArgGluGlyGluLysGlyLysAsnGlyLysAspSerSer 98
Db      198 -----ACTGAGCGGCTCTATACAGAGAAAGAAAGAAATGATAGTGAT----- 239
QY      99 AlaGlyLysPheProProProAlaGlnValGluProThrProGluAlaGluLeu 118
Db      239 ----- 239
QY      119 AlaGlnProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaIaGlyGlu 138
Db      240 -----AGAAAAGAGCAACCAAGTGGGT-----GCAACAAATGGGAG 278
QY      139 GluGluTrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgProSer 158

```



QY 223 GlyLeuTyrSerIysArgAlaAlaIysSerAsp-----AspThr 236  
Db 773 -----ATGCAATTTCACCTGCGCAAGCGTCAGTGAGGTTGCAGAGATCGACCA 826  
QY 237 SerAspAspPheMetGluGly 245  
Db 827 GACCCAGACCCAGACACTGCGAGATCGA 853  
RESULT 14  
AAK51987  
ID AAK51987 standard; cDNA; 7516 BP.  
AAK51987:  
XX 06-NOV-2001 (first entry)  
XX Human polynucleotide SEQ ID NO 532.  
DE Human polynucleotide SEQ ID NO 532.  
XX  
XX Human: cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200157190-A2.  
XX 09-AUG-2001.  
XX  
XX 05-FEB-2001; 2001WO-US04098.  
XX  
XX 03-FEB-2000; 2000US-0496914.  
XX 27-APR-2000; 2000US-0560875.  
XX 20-JUN-2000; 2000US-0598075.  
XX 19-JUL-2000; 2000US-0620325.  
XX 01-SEP-2000; 2000US-0654936.  
XX 15-SEP-2000; 2000US-0663561.  
XX 20-OCT-2000; 2000US-0693325.  
XX 30-NOV-2000; 2000US-0728422.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX  
XX WPI: 2001-476283/51.  
XX P-PSDB; AAM78854.  
XX  
XX Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -  
XX  
XX Claim 1: Page 1918-1926; 6221pp; English.  
XX  
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM7833-AAM80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activity/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.  
XX  
XX Sequence 7516 BP; 1909 A; 2056 C; 2326 G; 1225 T; 0 other:  
XX  
XX Alignment Scores: 1.49 Length: 7516  
XX Pred. No.:  
XX

Score: 162.00 Matches: 99  
Percent Similarity: 35.73% Conservative: 65  
Best local Similarity: 21.57% Mismatches: 183  
Query Match: 8.48% Indels: 113  
DB: 22 Gaps: 15  
US-09-972-758a-2 (1-359) x AAK51987 (1-7516)  
QY 2 AlagluProPheLeuSerGluTyrGln-----HisGlnProGln--- 14  
Db 2202 GCTGATGCTACCGCTGAGAGACAGAACCCCACTTTGTCCGCTGATCTATCCCAACCA 2261  
QY 15 -----ThrSerAsnCysThrGlyAlaAlaIaValGlnGlu 26  
Db 2262 CGAAGAAAGAGCGCGCAGCTGGACCCGCACTCTCGTCTGAGACACCTGCGTGCAGAG 2321  
QY 27 -----GluLeuAsnProGluArgPro-----ProGly 35  
Db 2322 TGTTCGAGAGCGCATCCGATCTGCGCGCAGGCGTTCCCAACAGAGGTGGCTTTCAGCA 2381  
QY 36 AlagluGluArgValProGluGluAspSerArgTrpGlnSerArgAlaPhe-----Pro 53  
Db 2382 GTTTCGCGAGAGATATGAGATCTGACTCCAACTCCATTCACAGAGGTTTCATGGACGG 2441  
QY 54 GlnLeuGlyGlyArgProGlyProGluGlyGlySerLeuGluSerGlnProProPro 73  
Db 2442 GAAGCAGAGCGCGTGGTGTATGATMAAGCCCTGGAGAGCTGCAGACGATCTGATCCGAT 2501  
QY 74 LeuGlnThrGlnAlaCysProGluSerSerCys----- 84  
Db 2502 TGGCCAGAGCAAAAGTCTTCTTCCTGCGCGTGTGCGCCACCTGCAGAGAGCGAGAGA 2561  
QY 85 -----LeuArgGluGlyGlyGln 92  
Db 2562 CTTGAAGATATCCGACGCTATCATAGGTTTCCAGGCGTGTGCAGAGGCTACTTGGCCAG 2621  
QY 93 AsnGlyAspAspSerSerAlaGlyIysPhe-----ProProAlaGluVal 109  
Db 2622 GAAGCATTTGCCAAGCGGCGAGCAGCAGCTTACCGCATGAGAGTCTCCAGCGAAGCT 2681  
QY 110 GlnProThrProGluAlaGluLeuAla-----GlnProCysHisAspSerGluAla 127  
Db 2682 CGCTGCCCTTACCTAAGCTGCGGAACTGCGAGCTGCGCGCTCTTCAACAGCTCAAGCC 2741  
QY 128 SerLeuGluAlaProAlaIaGlyGluGluGluTrp-----GlyGln 143  
Db 2742 GCTGCTGCAGAGTACCGCGAGAGAGATGATGATGCCAAGAGAGAGAGCTTGATGA 2801  
QY 144 GlnGlnArgGln-----LeuGlyLysLysHisArgArgArgProSerLysLys 160  
Db 2802 GGTACAGAGAAACACACTGGCTGCGGAGAAACAGCTCACGGAATGAGAGACUCCTGCAGTC 2861  
QY 161 LysArgHisTrp-LysProTyrTyrLysLeuThrTrpGluGluLysLysPheAspG 180  
Db 2862 TCAGCTCATGCGCAGAG-----AAATTCAGCTGCAGAGACACTCCAGCGCAAGAAG 2912  
QY 180 uLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaIaIaLysGlnPro 200  
Db 2913 CGAGCTGTGTGCGAGAGCTGAGAGCTCGGCGCGCTGACCGCCCAAGAGCAGGAATT 2972  
QY 200 AlaPro----- 202  
Db 2973 AGAAGAGATCTGCCATGACCTAGAGCGCAGGCTGGAGAGAGAGAGAGAGCGTGCACGA 3032  
QY 203 -----TyrAsnThrTrpGlnPheLeuMetAspAspHisAs 214  
Db 3033 CCTGCAGCGCGAGAGAGAAGAATGCGAGCAAGACATCCAGAGCTTGAGAGAGACCTGGA 3092  
QY 214 pGlnGluGluProAspLeuLysThrGlyLeuTyrSerLysArgAlaAlaIaLysSerAs 234  
Db 3093 GAGAGAGAG-----AGCGCGCGCGAGAAAGCT-GCAGCTGAGAGAAAGGTA 3136  
QY 234 pAspThrSerAspAspPheMetGluGluGlyGlyGluAspGlyLysSerAspG 254

```

Db      3137 CCACCGAGCGGAGCTGAAAAAACCTGGAGAGGAGGAGATCATCTGGAGAGGAGACT 3196
Oy      254 ymetgllyglYAspGlySerGluPheLeuGlnArgAspPheSerGluThrArgGluArgTy 274
Db      3197 GCAGCTGGCGCAAGAAAGAACTGCTGGAAGACAGAAATAGCTGACTTACCAACCAAC 3256
Oy      274 rHis-ThrGluSerLeuGlnAsnMetSerLysGlnGluLeuIleLysGluTyr----- 291
Db      3257 TCACAGAGAGGAGAGAAATTAAGAGCTCCCAAGCTCCAGAAACCAAGCATGAGGCCAA 3316
Oy      292 -----LeuGluLeuGluLysCysLeuSerArgMetGluAspGluAsnAsnArgLeuArgL 310
Db      3317 TGATCACTGCTTGGAGAGCGCTCCGCGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3376
Oy      310 euGluSerLysArgLeuGlyLysP-----A 319
Db      3377 AGACCGCGCGGAGAGCTGGAGGAGACTCCACAGACTCAGCGAGCAGACAGATCGCGAGCTCC 3436
Oy      319 sPAlaArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnL 339
Db      3437 AGCGCCAGATCGCGAGAGCTCAAGATGAGCTGGCGCAAGAAAGAGGAGGAGCTCCAGGCC 3496
Oy      339 euLeuThrGluAsnGluLeuHisArgGlnGlnGlnArgAlaProLeuSerLys 356
Db      3497 CCTGTGGCAGAGTGTGAAGAGAGAGCTGCCAGAAAGAACATGGCCCTCAAGAGAG 3549

RESULT 15
AB054663
ID. AB054663 standard; cDNA; 4041 BP.
XX
AC AB054663;
XX
DT 22-AUG-2002 (first entry)
XX
DE Human ovarian antigen HCOMB65 cDNA, SEQ ID NO:543.
XX
DE Human ovarian antigen; ovary; ovarian; breast; cancer; tumour;
XX
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
XX
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
XX
KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
XX
KW inflammatory condition; immune disorder; blood disorder;
XX
KW cardiovascular disorder; respiratory disorder; neurological disorder;
XX
KW gastrointestinal disorder; urinary system disorder; drug screening;
XX
KW gene therapy; chromosome mapping; forensic analysis;
XX
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
XX
KW antiinflammatory; gynaecological; reproductive; chromosome 22q13.1;
XX
KW gene; ss.
XX
OS Homo sapiens.
XX
PN WO200200677-A1.
XX
PD 03-JAN-2002.
XX
PF 07-JUN-2001; 2001WO-US18569.
XX
PR 07-JUN-2000; 2000US-209467P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Birse CE, Rosen CA;
XX
DR MPI: 2002-147878/19.
XX
DR P-PSDB: ABP41586.
XX
PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
XX
PT useful in the prevention, treatment and diagnosis of cancer (e.g.
XX
PT ovarian cancer), immune disorders, cardiovascular disorders and
XX
PT neurological diseases -
XX
PS Claim 1; SEQ ID NO 543; 2922bp; English.

```

```

CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC APP43228) and to cDNAs encoding them (AB054131-AB056305), and also
CC encompasses polypeptides 90% identical and polynucleotides 95% identical
CC to the sequences of the invention. The invention additionally relates to
CC recombinant vectors and host cells comprising human ovarian antigen
CC polynucleotides, antibodies against human ovarian antigens, and the use
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC treating, prognosing or preventing various ovary and/or breast-related
CC disorders. Such conditions include ovarian cancer and breast cancer, and
CC metastatic tumours of ovarian or breast origin, reproductive system
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC vaginitis), immune disorders (e.g., congenital and acquired
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC respiratory disorders, neurological disorders, gastrointestinal disorders
CC and urinary system disorders. Ovarian antigen polypeptides and
CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents cDNA encoding a human ovarian antigen of the
CC invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 4041 BP; 1039 A; 1092 C; 1259 G; 643 T; 8 other;

```

## Alignment Scores:

```

Pred. No.: 1 12 Length: 4041
Score: 159.50 Matches: 117
Percent Similarity: 32.34% Conservative: 46
Best Local Similarity: 23.21% Mismatches: 144
Query Match: 8.35% Indels: 198
DB: 24 Gaps: 20

```

US-09-972-758a-2 (1-359) x AB054663 (1-4041)

```

Oy      3 GluProPheLeuSerGluThrGlnHisGlnProGlnThrSerAsnGlyThrGlyAlaAla 22
Db      682 GAACCGGAGAGAGCTGAGCTGAGCAGCAAGCTCA-----GCAGG 723
Oy      23 AlaValGlnGlnGluLeuAsnPro-----GluArgPro 33
Db      724 GGAGGAGAGAGAAAGAAATTCCTCCGGAGAGCAGCTGGAGAGAGAGAGGCCAAGCA 783
Oy      34 ProGlyAlaGluGluArgValProGluGluAspSerArgTrpIleSerArgAlaPhePro 53
Db      784 CCGGAGAGAGCAGATGCCACCT-----CCA 810
Oy      54 GlnLeuGlyGlyArg-----ProGly----- 60
Db      811 TGCCAGGTGGCCGACATGAAAGAAAGATGAGAGCAGTGTGGGTGCTGGAATCTGC 870
Oy      61 -----ProGluGlyGlyGlySerLeuGlnSerGlnProPro 73
Db      871 TGAGAGGTGAAGAGAGAGAGCTCCAGAGACCTGGAGG-----CCTGAGCCA 918
Oy      74 LeuGlnThrGlnAlaCysProGluSerSerCysLeuArgGluGlyLysGlyGlnAsn 93
Db      919 -----GGCGCAGAGAGAGAGGTGGCCGCTTACGACAAAGCTGGAGAGACCAAGAC 969
Oy      94 -----GlyAspAspSerSerAlaGlyLysAspPheProPro--ProAlaGlu 108
Db      970 GCGGCTGACAGCAGAGACTGTGAGACCTGCGTGGTGGACCTGGACACACAGGCGCAGGCC 1029
Oy      109 ValGluProThrProGluAlaGlu-----Leu 117

```

```
Db 1030 GTGCAACCTGAGAGAACGAGAAAGATTGTACCAGCTCTCGCGGAGAGAACCAAT 1089
OY 118 LeuAlaGlnProCysHisAspSerGluAlaSerLysLeuGlyAlaPro----- 133
    ||| ||| |||
Db 1090 CTCTGCGCAAGTATGACAGAGAGCCGACCGGGCTGAGCGGAGGCCCCAGAGAAAGAGAC 1149
OY 134 -----AlaAlaGlyGlyGlu 139
    ||| ||| |||
Db 1150 CAAGGCTCTGTCGGTGGCCCGGCGCTGAGAGAACCCATGAGCGCAAGAGCGGACTGAG 1209
OY 140 GluTrpGlyGlnGlnArg-----GlnLeu 148
    ::||| |||
Db 1210 CGCCTCAACCAAGCATTCGCCAGCAGATGCGACCTTATGACCTCCACAGATGATGTG 1269
OY 149 GlyLysLysLysHisArgArgArgProSerLysLysLysArgHisTrpLysProTyrTyr 168
    ||||| |||
Db 1270 GGCAGAGAGTGTCCAGACAGCTGGACAGATCCAAAGCGGCCCTAGACAGCAGGTGAGAGAG 1329
OY 169 LysLeuThrTrpGluGluLysLysLysPheAspGluLysGlnSerLeuArgAlaSerArg 188
    ||| |||
Db 1330 ATGAAGACGCGAGCTGCAAGAGCTGAG--GACGAGCTGCGAGGCCACCGAGATGCCAAG 1386
OY 189 IleArgAlaGluMetPheAlaLysGlyLysProValAlaProTyrAsnThrThrGlnPhe 208
    ::||| |||
Db 1387 CTGCGGTTGGAGGTCAACCTGCGAGCCATGAGGCC----- 1422
OY 209 LeuMetAspAspHisAspGluGluGluProAspLeu-LysThrGlyLeuTyrSerLysAr 228
    ||| ||| ||||| ::|
Db 1423 -----CACTTGCAGCGGAGACCTGCGAGGCCGCGGACGACAGACCGAG 1464
OY 228 GAlaAlaAlaLysSer-AspAspThrSerAspAspPheMetGluGluGlyGlyGlu 248
    ||| ||| |||
Db 1465 GAGAAAGAAAGACAGCTGTCACAGAGTGCAGAGATGAGGAGCGACAGCTGAGAGCGAG 1524
OY 248 LysAspGlyGlySerAspGlyMetGly-----GlyAspGlySerGlu-Phe 262
    ||| ||||| |||
Db 1525 AGGAGAGCAGCGCTGATGGCAGTGGCCCGCAGAGAAAGCTGAGATGACCTGAGAGCAC 1584
OY 263 LeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGlnAsnMet 282
    |||||
Db 1585 CTGAGAGCGGCACATC-----GACTCGGCCAACAGAACCGGAGAGAGCATCAACAGCTG 1641
OY 283 -----Ser 283
    |||
Db 1642 CGGAAGCTGCAGGCCCAAGATGACAGACTGCATGCCGAGCTGATGACACCCGCGCTCT 1701
OY 284 LysGlnGlnLeuIleLysGluTyrLeuGlnLeuGlnLysCysLeuSerArgMetGluAsp 303
    ::||| ||| ||| |||||
Db 1702 CGTGAAGAGATCTGGGCCCAAGCCAAAGAGAGAAAGAAAGCTGAAGAGCATGAGAGGCC 1761
OY 304 GluAsnAsnArgLeuArgLeuGluSer----- 312
    ||| ::||| |||
Db 1762 GAGATGATCCAGTTGCAAGAGAACTGGCAGCCCGGAGCTGCCAAGCGCCAGGCCAGC 1821
OY 312 ----- 312
    |||
Db 1822 AGGAGCGGATGAGCTGGCTGACGAGATCCGCAACAGAGCGGCAAGAGCCCTGGCT 1881
OY 313 ---LysArgLeuGlnGlyLysAspAlaArgValArgGlnLeuGlnLeuGlnLeu----- 329
    ::||| ||| ::||| ||| |||||
Db 1882 TAGAGGAGAGCGGCTGTG-GAGGCCCGCATCGCCAGCTGAGAGAGAGCTGAGAGAG 1940
OY 330 -----AspArgLeuArgAlaGlnLeuGlnLeu--- 339
    ||||| |||
Db 1941 GAGCAGGGCAACAGGAGCTATGACAGACCGGCTGAAGAAAGGCCAATGACATCGAC 2000
OY 340 -----LeuThrGluAsnGlnLeuHisArg 347
    ::||| |||
Db 2001 CAGATCAACACCGACTGAACCTGGAGCGGAGCCAGCGCCGAGAAAGAGAAATGCTGG 2060
OY 348 GlnGln 349
    |||||
Db 2061 CAGCAG 2066
```

Search completed: June 17, 2003, 10:54:16  
Job time : 292 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: June 17, 2003, 09:52:47 ; Search time 60 Seconds  
(without alignments)  
1834.949 Million cell updates/sec

Title: US-09-972-758A-2

Perfect score: 1910  
Sequence: 1 MAFPLSEVQHOPQTSNCTG.....LTENELHROOERAPLSKFGD 359

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09972758/runatc\_17062003\_095238\_24394/app\_query.fasta\_1.519  
-DB=Issued Patents\_NA -OFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USPR=US09972758\_@CGN\_1\_1\_40\_@runatc\_17062003\_095238\_24394 -MCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONLOG  
-DEV\_TIMEOUT=120 -MARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6  
-FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents\_NA:\*  
1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*  
4: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*  
5: /cgn2\_6/prodata/1/ina/PTUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	158.5	8.3	9551	1	US-08-056-200-93 Sequence 93, Appl
2	158.5	8.3	9551	2	US-08-800-644-93 Sequence 93, Appl
3	155	8.1	2277	1	US-08-676-967-5 Sequence 5, Appl
4	155	8.1	2277	1	US-08-676-967-5 Sequence 5, Appl
5	155	8.1	2277	2	US-09-098-487-5 Sequence 5, Appl
6	141	7.4	2304	1	US-08-464-266-1 Sequence 1, Appl
7	141	7.4	2304	1	US-08-464-272-1 Sequence 1, Appl
8	141	7.4	2304	4	US-08-464-514-1 Sequence 1, Appl
9	141	7.4	2304	4	US-08-486-403-1 Sequence 1, Appl
10	137	7.2	1617	4	US-09-265-013-2 Sequence 2, Appl
11	137	7.2	7308	4	US-09-011-745-3 Sequence 3, Appl
12	137	7.2	7308	4	US-09-011-745-4 Sequence 4, Appl

13	137	7.2	7616	4	US-09-011-745-2	Sequence 2, Appl
14	137	7.2	8202	1	US-08-258-420-13	Sequence 13, Appl
15	137	7.2	8332	3	US-08-850-961-1	Sequence 1, Appl
16	137	7.2	8332	4	US-09-479-776-1	Sequence 1, Appl
17	137	7.2	8332	4	US-09-309-572-11	Sequence 11, Appl
18	137	7.2	8332	4	US-09-315-127-1	Sequence 1, Appl
19	137	7.2	8332	4	US-09-265-013-1	Sequence 1, Appl
20	134	7.0	3930	3	US-09-162-373-2	Sequence 2, Appl
21	134	7.0	3930	4	US-09-467-946-2	Sequence 2, Appl
22	133.5	7.0	6755	3	US-08-931-999-4	Sequence 2, Appl
23	131.5	6.9	5661	4	US-08-938-105-2	Sequence 4, Appl
24	131	6.9	3023	4	US-09-593-589-10	Sequence 10, Appl
25	128.5	6.7	2581	4	US-09-370-838-66	Sequence 66, Appl
26	126	6.6	7791	3	US-08-949-386-23	Sequence 23, Appl
27	126	6.6	7791	3	US-08-450-562-23	Sequence 23, Appl
28	126	6.6	7791	4	US-08-984-709A-23	Sequence 23, Appl
29	126	6.6	7791	4	US-08-450-272-23	Sequence 23, Appl
30	126	6.6	7808	2	US-08-149-097D-22	Sequence 22, Appl
31	126	6.6	7808	3	US-08-949-386-22	Sequence 22, Appl
32	126	6.6	7808	3	US-08-450-562-22	Sequence 22, Appl
33	126	6.6	7808	4	US-08-984-709A-22	Sequence 22, Appl
34	126	6.6	7808	4	US-08-450-272-22	Sequence 22, Appl
35	125.5	6.6	2384	1	US-07-814-964-10	Sequence 10, Appl
36	125.5	6.6	2384	1	US-08-328-809-5	Sequence 10, Appl
37	125.5	6.6	2384	1	US-08-328-809-5	Sequence 10, Appl
38	125.5	6.6	2384	5	PCT-US92-11107-10	Sequence 10, Appl
39	125	6.5	2672	3	US-09-214-564A-5	Sequence 5, Appl
40	125	6.5	2815	3	US-09-214-564A-1	Sequence 1, Appl
41	124.5	6.5	1276	4	US-09-177-325-2	Sequence 2, Appl
42	124.5	6.5	1276	4	US-09-411-812A-2	Sequence 2, Appl
43	124.5	6.5	1276	4	US-09-590-113-2	Sequence 2, Appl
44	124	6.5	4722	4	US-08-979-608A-14	Sequence 14, Appl
45	122	6.4	3201	3	US-08-556-419-11	Sequence 11, Appl

#### ALIGNMENTS

RESULT 1  
US-08-056-200-93  
Sequence 93, Application US/08056200  
Patent No. 5616500  
GENERAL INFORMATION:  
APPLICANT: Steinert, Peter M.  
APPLICANT: Lee, Seung-Chul  
APPLICANT: Kim, In-Gyu  
APPLICANT: Chung, Soo-Il  
APPLICANT: Park, Sang-Chul  
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
NUMBER OF SEQUENCES: 117  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/056,200  
FILING DATE: 30-APR-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Fedrick, Michael F.  
REGISTRATION NUMBER: 36,799  
REFERENCE/DOCKET NUMBER: NIH054.001A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (714) 760-0404  
TELEFAX: (714) 760-9502





TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1507..1644  
 FEATURE:  
 NAME/KEY: Intron  
 LOCATION: 1645..2511  
 NAME/KEY: CDS  
 LOCATION: 2512..8070  
 US-08-800-644-93

## Alignment Scores:

Pred. No.: 0 00154  
 Score: 158.50  
 Percent Similarity: 39.89%  
 Best Local Similarity: 25.56%  
 Query Match: 8.30%  
 DB: 2  
 Length: 9551  
 Matches: 91  
 Conservative: 51  
 Mismatches: 135  
 Indels: 80  
 Gaps: 16

US-09-972-758a-2 (1-359) x US-08-800-644-93 (1-9551)

QY 31 GUAArgProProGlyAlaGluGluArg-----ValProGluGlu 43  
 Db 4810 GAGAGCGCAGCTGAGCGCGCAGAGCGCCAGCAGCGGAGACACCGCTTCTCCGCGAGAG 4869  
 QY 44 AspSerArGTrpGlnSerArGAlaPheProGlnLeuGlyGlyArGProGlyProGluGly 63  
 Db 4870 GAGGAGAGAGCAGCGCC-----GGCCCGCAGCGCAGAGAGG 4908  
 QY 64 GlnGlySerLeuGlnSerGlnProProProLeuGlnThrGlnAlaCysProGlnSer 83  
 Db 4909 GAGAAAGAGCTGCAGTTCTCGAGCAAGAGAGAGCAGCTCCAGCGCGGAGCGGAGCCCAA 4968  
 QY 84 CysLeuArgGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 102  
 Db 4969 CAGCTCCAG 5027  
 QY 103 PhePro---ProProAlaGluVal-----GluProThrProGlu 114  
 Db 5028 GCGCGCGAGCCAAATGAGAGTGGCACTGAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAG 5087  
 QY 115 AlaGluLeuLeuAlaGlnProCysHisAspSerGlnAlaSerGlyAlaProAla 134  
 Db 5088 GCTGTACGCCAAGCCACCTACAAAGACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5141  
 QY 135 AlaGlyGly-GlnGluGlnTrpGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 154  
 Db 5142 GCAG 5183  
 QY 154 GArGArProSerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 174  
 Db 5184 GAGAAAGCGCCCAAG 5243  
 QY 174 uLysLysLysPheAspGlnLysGlnSerLeuArgAlaSerArgGlnLeuArgAlaGlnMetPh 194  
 Db 5244 AGAGCAGCTGCTGAG 5303  
 QY 194 eAlaLysGlyGlnProValAlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAs 214  
 Db 5304 TCGGAGAGATTAAGAG-----CTGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5336  
 QY 214 pGln-----GlnGluProAspLeuLysThrGlyLeuTyrSerLysArgAlaAlaAl 231  
 Db 5337 GCACCTCTGAG 5381  
 QY 231 aLysSerAspAspThrSerAspAspAspPheMetGlnGlyGlyGlyGlyGlyGlyGlyGlyGly 251  
 Db 5382 GGAAGAAAATACCGCAG 5421

QY 251 ySerAspGlyMetGlyLysAspGlySerGlnPheLeuGlnArgAspPheSerGlnThrTr 271  
 Db 5422 -----GAAGAGCAGACTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5456  
 QY 271 rGluArg-----TyrHisThrGlnSerLeuGlnAsnMetSerLysG 285  
 Db 5457 AAGCGCCAGAGCTGGAG 5516  
 QY 285 nGlnLeuLleLysGlnGlyTrLeuGlnGlyLysCysLeuSerArGmetGluAsp----- 303  
 Db 5517 GCAGCTGCTGAG 5576  
 QY 304 -GluAsnAsnArgLeuArgLeuGlnSerLysArgGlyGlyAspAspAlaArgValAr 323  
 Db 5577 GGAG 5636  
 QY 323 g-----GlnGluGlnLeuGlnLeu-----AspArgLeuArgAlaGlnAsnLe 337  
 Db 5637 AAGCGCCAGAGAGCTGGAG 5696  
 QY 337 uGlnLeuLeuThrGlnAsn---GlnLeuHisArgGlnGlnGluArg 351  
 Db 5697 GCAGCTGCTGAG 5742

## RESULT 3

US-08-676-967-5

Sequence 5, Application US/08676967

Patent No. 5747317

GENERAL INFORMATION:

APPLICANT: COLLINS, KATHLEEN

TITLE OF INVENTION: Human Telomerase

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESS: Science & Technology Law Group

STREET: 268 Bush Street, Suite 3200

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/676,967

FILING DATE:

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Osman Ph.D., Richard A

REGISTRATION NUMBER: 36,627

REFERENCE/DOCKET NUMBER: UCB96-055

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415)343-4341

TELEFAX: (415)343-4342

INFORMATION FOR SEO ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 2277 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-676-967-5

## Alignment Scores:

Pred. No.: 0.000421  
 Score: 155.00  
 Percent Similarity: 33.93%  
 Best Local Similarity: 21.57%  
 Query Match: 8.12%  
 DB: 1  
 Length: 2277  
 Matches: 96  
 Conservative: 55  
 Mismatches: 147  
 Indels: 148  
 Gaps: 18

US-09-972-758a-2 (1-359) x US-08-676-967-5 (1-2277)

```

OY      6 LeuSerGIuTYrGlnHISGlnProGlnThrSerAsnCysThrGlyAlaAlaValaGln 25
      144 CTAGTGACCTTCAGCAT-----GCTGGAGAGACGTGCACCGCGC 182
OY      26 GlnGlu-----LeuAsnProGluArgProProGlyAlaGln 38
      183 CCTGAGAGAGATCACACCTTCGAGGCTGCAGATGACCTGACCGGCGCAAGAA 242
OY      39 ArgValProGlnGluAspSerArgTrpInSerArg-----AlaPheProGlnLeuGly 56
      243 GCTGGCGACAGACAGCAAGAGAAAGGCAAGACAGACAGCAGCTGCCCAAGAA 302
OY      57 GlnArgProGlyProGlnGlyGlnGlySerLeuGlnSerGlnPro----- 71
      303 GCCCAA---GCCCAAGAGAGCGCAAGGTGCGCGACAAAGAGCGCGCTGATCATCCGCA 359
OY      72 ProProLeuGlnThrGlnAlaCysProGlnSerCysLeuArgGlnGlyGlnGly 91
      360 CCTGAGCTTCAAGTGCACCGGAGACGACCTGAGACCGCTGTCGCCAGTTCGGGCGC--- 416
OY      92 GlnAsnGlyAspAspSerSerAlaGlyGlyAspPheProProAlaGlnValaGlnPro 111
      417 -----CCTGCTGGAGGTGACATCCCGCGCAAGCCGACGCGCA--- 455
OY      112 ThrProGlnAlaGlnLeuLeuAlaGlnProCysHisAspSerGlnAlaSerLysLeuGly 131
      456 -----GATGCGCGCTTCGGCTTCGTGCAGTTCAGAACTGCTGAGCGCGCGCAAGC 509
OY      132 AlaProAlaAlaGlyGlyGlnGlu----- 140
      510 -----CCTGAAGGCATGAACATGAAGAGATCAAGGCGCGCACCGCTGGAGCTG 563
OY      141 -----TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysHisArgArg-ArgPr 157
      564 GCGCGTGGCCAGGAGCAAGTACAAAGACACCCAGAGCTGAGCGCATCGCGAGAGAA 623
OY      157 oSerLysLysLysArgHisTrpLysProTyTrpLysLeuThrTrpGlnGlnLysLys 177
      624 GAGCGACAGAGACAGCACCGAGAGCGTCAAGAAAGCGCGCGAGAGAGACAT 683
OY      177 sPheAspGlnLysGlnSerLeuArgAlaSerArgIleArgAlaGlnMetPheAlaLysGln 197
      684 GGAGAGAGAGAGAAACGAC----- 702
OY      197 yGlnProValAlaProTyAsnThrThrGlnPheLeuMetAspAspHisAspGlnGln 217
      703 -GAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCGGCTGTCAGCAGCAGCAGCAGCAG 761
OY      217 uProAspLeuLysThrGlyLeu-----TyrSerLysArgAlaAla---- 230
      762 GGAGACATCGAGCAGCAGGTGACCAAGCCCGTCAGATCCAGAAAGCGCGCGTGAAGCG 821
OY      231 -----AlaLysSerAspAspThrSerAspAspAspPhe---MetGlnGlnGly 246
      822 CCCCCCGCCCGCAAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 881
OY      246 yGlnGluAspGlySerAspGlyMetGlyGlyAspGlySerGlnPheLeuGlnArgAs 266
      882 CAGCATGAC-----GACCGCAGAGAGCTGGCGCCAGAGCGA 917
OY      266 pPheSerGlnThrTyTrpGlnArgTyHisThrGlnSerLeuGlnAsnMetSerLysGln-- 285
      918 CACACGACACCGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 977
OY      285 ----- 285
      978 GCTGCCAGCAGCAGTGAACAGAGCGCAAGCCGTTTCATCCGCAACCTGACCTTCAGACG 1037
OY      286 -----GlnLeuIleLysGlnTyTrpLeuGln----- 294
      1038 CGAGAGAGAGAGCTGGCGGAGCTGCTGACAGCAGTTCGGCGAGCTGAAGTACGTGCGCAT 1097

```

```

OY      294 ----- 294
Db      1098 CGTGTGACCCCGACACCGAGCAGACAGAGGCTGGCGCTTCGCCCATCATGACCCA 1157
OY      295 -----GlnLysCysLeu-----SerArgMetGlnAspGlnAsnArgLeuArg 309
      1158 GGAGCGCGCCCGCAGAAAGTCTGCTGGCGCGCGAGCAGCAGCAGCAGCAGCAGCAG 1217
OY      309 gLeuGlnSerLysArgLeuGlnGlyAspAspAlaArgValaArgGlnLeuGlnGln 329
      1218 GCTGGACCGCGCCAGCTG-----AAGGTGGACCTGGCGCGT 1253
OY      329 uAspArgLeuArgAlaGlnAsnLeuGlnLeuThrGlnAsnGlnLeuHisArgGlnGln 349
      1254 GACCGCGCAGCAGCGCGCGCAGCTGACAGACCAAGGT-CAAGAAAGCGCACCGGCAACC 1312
OY      349 nGluArgAlaPro 353
      1313 GCAACCTGTACT 1325
Db

RESULT 4
US-08-676-974-5
: Sequence 5, Application US/08676974
: Patent No. 5770422
: GENERAL INFORMATION:
: APPLICANT: COLLINS, KATHLEEN
: TITLE OF INVENTION: Human Telomerase
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Science & Technology Law Group
: STREET: 268 Bush Street, Suite 3200
: CITY: San Francisco
: STATE: CA
: COUNTRY: USA
: ZIP: 94104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/676,974
: FILING DATE:
: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: Osman Ph.D., Richard A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: UCB96-055
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415)343-4341
: TELEFAX: (415)343-4342
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2277 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-676-974-5

Alignment Scores:
Pred. No.: 0.000421 Length: 2277
Score: 155.00 Matches: 96
Percent Similarity: 33.93% Conservative: 55
Best Local Similarity: 21.57% Mismatches: 147
Query Match: 8.12% Indels: 148
DB: 1 Gaps: 18

US-09-972-758A-2 (1-359) x US-08-676-974-5 (1-2277)
OY      6 LeuSerGIuTYrGlnHISGlnProGlnThrSerAsnCysThrGlyAlaAlaValaGln 25
      144 CTAGTGACCTTCAGCAT-----CCTGGAGAGACGTGCACCGCGC 182

```

OY		26	GIuGlU-----LeuAsnProGluArgProProGlyAlaGluLu	38
Db		183	CCTGAAGAGATCACCACCTTCGAGGGCTGCAGATCACTGACCGCCAAAGAA	242
OY		39	ArGValProGluGlnAspSerfArgPrpGlnSerArg-----AlaPheProGlnLeuGly	56
Db		243	CCTGCCAACAAGACCAGAAGGGCAAGACAGAACAGAACACCGACTGCCCAACAAGCA	302
OY		57	GlyArgProGlyProGluGlyGlnGlySerLeuGlnUserGlnPro-----	71
Db		303	GCCCA---GGCCAGAGAGGCCAACAGGTGGCCACAAAGAGCCCCGCTCATCATCCGCAA	359
OY		72	ProProLeuGlnInThGlnAlaAcysProGluSerCysSleuArgGlnGlyGlnLysGly	91
Db		360	CCTGAGCTTCAAGTGACAGGAGAGACGACCTGAGACCGTTGGTCCGCGGC---	416
OY		92	GlnAsnGlyAspAspSerSerAlaGlyGlyAspPheProProAlaGluValGluPro	111
Db		417	-----CGTGTGTGAGTGAACATCCCCCCAGACCCCGAAGCGCAA---	455
OY		112	ThrProGluAlaGlnLeuLeuAlaGlnProCysHisAspSerGlnAlaSerLysLeuGly	131
Db		456	-----GATGGCGCGGCTTCGGCTTCGTCCATTCAAGAACTGTGGAGGCGCAAGGC	509
OY		132	AlaProAlaIaagLyGlyGluGluLu-----	140
Db		510	-----CCTGAAGGGCATGAACATGTAAGAGATCAAGGGCCGACCGGTGGCTGACTG	563
OY		141	-----TPGlyGlnGlnGlnArgGlnLeuGlnGlyLysLysHisArgArg-ArgPr	157
Db		564	CGCCGTGCCCAAGGACAGATCAACAGACACCCAGACCGGTGAGCCCATTCGGCAGAGAA	623
OY		157	oSerLysLysArgHisTrpLysProTyrrTyrLysLeuThrTrpGluGlnLysLysLy	177
Db		624	GAGCCAGCAGAGACACAGACAGAGAGCGCTGAAGAAAGGCGCGCAGAGAGAGACAT	683
OY		177	sPheaSpLiuLysGlnSerLeuArgAlaSerfArgIleArgAlaGluMetPheAlaLysG	197
Db		684	CGAGAGAGAGAGAAACAC-----	702
OY		197	yGlnProValAlaProTyrrAsnThrThcGlnPheLeuMetLaspasPhisAspGlnGlu	217
Db		703	-GACAGCAGCACGACGACGACGACGAGGACGCGCTGTTCACAGCAGGAGAGAGAGA	761
OY		217	uProAspLeuLysThrGlyLeu-----TyrSerLysArgAlaIaIa-----	230
Db		762	GGAGAACTCGAGAGACAGCAAGGTGACCAACCCGCTGCAGATCCAGAAGCGCGCTGACACG	821
OY		231	-----AlaLysSerAspAspTrhSerAspAspPhe--MetGlnGluGlyG	246
Db		822	CCCCCCCCCCCAGAGACGACGACACACAGAGAGAGAGACGACGACCTCGAGAGAGAGA	881
OY		246	yGluGlnAspGlyGlySerAspGlyMetGlyGlyAspGlySerGluPheLeuGlnArgAs	266
Db		882	CAGACATGCAC-----GACGGCAGAGAGCTGCGCCCAAGAGCA	917
OY		266	pPheSerGluThrTyrrGluArgTyrrHisThcLuserLeuGlnAsnMetSerLysGln-	285
Db		918	CACCAAGCACCGAGAGCAGAGAGACAGACGCGCTGCAGTGAACAMCAGAAAGAAAGCGCA	977
OY		285	-----	285
Db		978	GCTGCCAGCAGCAGTGAACGAGCGCAGACCGTGTTCATCCGCAACTGACCTTCGACAG	1037
OY		286	-----GluLeuIleLysGluThrLeuGlnLeu-----	294
Db		1038	CGAGAGAGAGAGCTGGCGCAGCTGCTGCACAGTTCGCGCAGCTGAAGTAGTACGTCCGAT	1099
OY		294	-----	294
Db		1098	CGTGTGTCACCCGACACCGACAGACAGCAGAGGCTGCCTTCGCGCCAGTTTCATACCCA	1157

```

QY 295 -----GlibuScysLeu-----SerArgMetGluAspGluAsnAsnArgLeuAr 305
Db 1158 GGAAGCGCCGACAGAGAGTGGCTGCTGGGCCCCCAGCCCCCGAGAACGGAGCGCGGCGCTGAA 121
QY 309 GLeuGluSerLysArgLeuClYGLysAspSpsAlaArgValArgGluLeuGluLeuGluLe 325
Db 1218 GCTGGAGCGCGCGGACGCTG-----AAGGTGGACCTGGCCGT 125
QY 329 uASpArgLeuArgAlaGluAsnLeuGluLeuLeuThrGluAsnGluLeuHisArgGlnG 349
Db 1254 GACCCCGGAGAGAGCGCGCCAGAGTGCAGACCAACCAAGGT-GAAGAGACCCACACGGCACCC 131
QY 349 nGluArgAlaPro 353
Db 1313 GCAACCTGTACT 1325

RESULT 5
US-09-098-487-5
? Sequence 5, Application US/09098487
? Patent No. 5917025
? GENERAL INFORMATION:
? APPLICANT: COLLINS, Kathleen
? TITLE OF INVENTION: Human Telomerase
? NUMBER OF SEQUENCES: 11
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Science & Technology Law Group
? STREET: 268 Bush Street, Suite 3200
? CITY: San Francisco
? STATE: CA
? COUNTRY: USA
? ZIP: 94104
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/098,487
? FILING DATE:
? CLASSIFICATION:
? ATTORNEY/AGENT INFORMATION:
? NAME: Osman Ph.D., Richard A
? REGISTRATION NUMBER: 36,627
? REFERENCE/DOCKET NUMBER: UCB96-055
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (415)343-4341
? TELEFAX: (415)343-4342
? INFORMATION FOR SEQ ID NO: 5:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2277 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? US-09-098-487-5

Alignment Scores:
Pred. No.: 0.000421 Length: 2277
Score: 155.00 Matches: 96
Percent Similarity: 33.93% Conservative: 55
Best Local Similarity: 21.57% Mismatches: 147
Query Match: 8.12% Indels: 148
DB: 2 Gaps: 18
US-09-972-758A-2 (1-359) x US-09-098-487-5 (1-2277)
QY 6 LeuSerGluArgGlnHisGlnProGlnThrSerAsnGlyThrGlyAlaAlaValGln 25
Db 144 CTACGTGACCTTCAGCAT-----GCTGGAGGACACGTGCAGCGCGC 182
QY 26 GluGlu-----LeuAsnProGluArgGProProGluYAlaGluGlu 38
Db 183 CTCGAAGGAGATACCAACCTTCAGAGGCTGCAGATCAACGTGACCGTGCAGAGGAA 242

```



OY	13	ProGlnThrSerAsnGlyThrcylValAlaValGlnGlnGlnLeuAsnProGluArg	32
Db	388	CCCAATTGGCTGGAGGCTTCGCCCGCTGCAGTCCACACAGATATCCGCTTAACAT	447
OY	33	ProProGluAlaGluGluArg-----ValProGluGluAspSerArgTrpGlnSer	49
Db	448	CCGCTGACGGCAGACAGACACCTTCGCTCTATTTC--CGGGATCGGGCCAGTGGCAAGCA	506
OY	50	ArgAlaPheProGlnLeuGlyGlyArgProGly-----ProGluGlyGly	65
OY	507	CTACGGCTGTACACTGTGAGCGGTGCAGAGGGCTTTTAAAGCACAGTGGCAAGGA	566
OY	66	Ser-----LeuGlnSerGlnProProProLeuGlnThrGlnAlaGlySerGlnSer	82
Db	567	TCTCAGATACGGCTTCCAGGAGAAACCCAGCATCTCATAGACMAAGCGGCACAGCA	623
OY	83	SerGlySerLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGlyGlyAsp	102
Db	623	-----	623
OY	103	PheProProAlaGluValGlnProThrProGluAlaGluLeuLeuAlaGlnProGly	122
Db	624	-----CCGCTGCCAGTACTGCCCTACCGAGAGTCCCTAACCTGGG-----	665
OY	123	HisAspSerGlnAlaSerLysLeuGlyAla-----ProAlaAlaGlyGlyGlu	138
Db	666	CATGACGCGCAAGCGGTCCAGAGAGAGCGTCAACGCGCGCCCGCAATCGCGCGG	722
OY	139	GluGluTrpGlyGlnGlnGlnArg-----	146
Db	723	-----TAGGCTACGGCCAGCGGAGCGCGCCAGTACCGTCCGATGCGTAGCGCGAT	776
OY	147	GlnLeuGlyLysLysLysHisArgArgArgProSerLysLysArgHisTrpLysPro	166
Db	777	CAGCTCTCAAGCGCGGAGAGAGCAGCGG-----CGTTCTGGCGGAT	821
OY	167	TyrTrpLysLeuThrTrpGluGlnLysLysPheAspGlyLysGlnSerLeuArgAl	186
Db	822	-----GGGACGGCGGCAACCGTTCTGATGACTTCATATCCATATACCAATTAACCT	863
OY	186	AspArg-----LearGlnAlaGlnMetPheAlaLysGlyGln--	198
Db	864	GTCACGAGATTTCGATCATGAGCGCATCATAGAGCCGACGACGCGACACCAACATG	923
OY	199	-----ProValAlaProTyrAsnThrThrGlnPheLeu	210
Db	924	CGGGCATGTCGACTGACGATCTTCGCGGTGGTCCCTATTCCACAGTCCAG-----	975
OY	210	AspAspHisAspGlnGluProAspLeuLysThrGlyLeu-----	224
Db	976	-----CCGGACTACAGGGCTGCCGTCGTGCGCCCTGTGCCAAGT	1011
OY	225	-----TyrSerLysArgAlaAlaAlaLysSerAs	234
Db	1014	GGTCAACAAACAGCTCTTCCAGATGATGGAATATACGGCGCATGATGCGCACTTGGCCA	1077
OY	234	AspThrSerAspAspPhe-----	241
Db	1074	GGTGGCGCTGACGACGACGATGATTCGCTGAAGAACCCGCTTGATGACGCTCATTTGC	1133
OY	242	-----MetGlnGluGlyGlyGluGlnAspGlyGlySer	252
Db	1134	GAAGGTGGCTCGTGGACATATCGTTTGCCTGATATACGGCGST--GCCGGGGGGGGGG	1199
OY	252	AspGlyMetGlyGlyAspGlySer-----	260
Db	1191	CGGTGAGATAGGCCACGATGGCTCTTTGAGGAGACATACCGGCGCTTTCACGCCACGA	1250
OY	261	GlnPheLeuGlnArgAspPheSerGlyThrTyrrGluArgTrpTrpHisThrGlnSerLeu--	279
Db	1251	GCTGTCTCTCAACGAGACTTCTCG--TACCATTCGACAGATGGCAT	1291
OY	279	-----	279

Db	1296	CAAGCGCGGTGTGCACCCATCTTCGACCGCATATTGCGAGCTGATGTAAGATGAA	1355
QY	280	-----GIAsmeISeIrySGInGIuLeu-----	287
Db	1356	GCGGCTGAATTCGACCGACGCGAGTGTCTCTTGTAAGGCCATCACTGTACAACC	1415
QY	288	-----TLeysGIuTyrLeuGIuLeuGIuLylScysLeuSerArgmeGIu--	302
Db	1416	GGACATACGGCGGTACAAGACCGCGCGGAGATGCAATGATGCCCGAGAACGTACGC	1475
QY	303	-----ASpGIuAsnArgLeuArgLeuGIuSerLysArgLeuGIyGIyAspAspAl	320
Db	1476	TTTGCTTGGACGAGCATGCCCTCGGAACATCC-----GGCGAGATGG	1520
QY	320	aArGIuValArgGIuLeuGIuLeuGIuLeuAspArg-LeuArgAlaGIuAsnLeuLeuL	340
Db	1521	ACGGCTTGGCCAACTCTGCTGCGCTGCGCGCTTGGCATCATACGCTCGAAGTCC	1580
QY	340	eutHrGIuAsnGIuLeuHISArgGIuGIuGIuArgAlaProLeuSerLys	356
Db	1581	AGGATCACTTCTCTTCCGCAATTACACGACGCGGCCCTGTGAGAGAC	1650

```

: RESULT 7
: US-08-464-272-1
: Sequence 1, Application US/08464272
: Patent No. 5688691
: GENERAL INFORMATION:
: APPLICANT: ORO, Ph.D., ANTHONY E.
: APPLICANT: EVANS, Ph.D., RONALD M.
: TITLE OF INVENTION: INSECT RETINOID-LIKE RECEPTOR
: TITLE OF INVENTION: COMPOSITIONS AND METHODS
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
: STREET: 444 South Flower Street, Suite 2000
: CITY: Los Angeles
: STATE: CA
: COUNTRY: United States
: ZIP: 90071
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/464,272
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/013,975
: FILING DATE: 04-FEB-1993
: APPLICATION NUMBER: US 07/497,935
: FILING DATE: 22-FEB-1990
: ATTORNEY/AGENT INFORMATION:
: NAME: Reiter, Stephen E.
: REGISTRATION NUMBER: 31,192
: REFERENCE/DOCKET NUMBER: P41 9350
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 619-546-4737
: TELEFAX: 619-546-9392
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2304 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 163..1704
: US-08-464-272-1

```

Alignment Scores:

Pred. No.:	0.00669	Length:	230
Score:	141.00	Matches:	16
Percent Similarity:	32.43%	Conservative:	406
Best Local Similarity:	22.18%	Mismatches:	127
Query Match:	7.38%	Indels:	197
DB:	1	Gaps:	22

US-09-972-758A-2 (1-359) x US-08-464-272-1 (1-2304)

QY	13	ProGlnIrrSerAsnCySThgIyAlaAlaAlaValGlnGlnGluIleuAsnProGluArg	32
Db	388	CCCAATTGGCTGGAGGCTGTGCGCGCCCTGACGTCAGCCAGACAGACATATCCGCTTAACCT	447
QY	33	ProProGluAlaGluLysArg-----ValProGluGluAspSerArgTrpGlnSer	49
Db	448	CCGCTGAGCGGACAGACACACCTTCGCTCTATTTC--CGAGGATCGGCGACAGTGGCAGACA	506
QY	50	ArgAlaPheProGlnLeuGlyGlyArgProGly-----ProGluGlyGlyGly	65
Db	507	CTACGGCGCTGTACAGCTGTGAGGGGTGCAGAGGCTTCTTAAACGCACAGTCCGCAGAGA	566
QY	66	Ser-----LeuGluSerGlnProProProLeuGlnIrrHAlaCysProGluSer	82
Db	567	TCTCAATACAGCTTTCAGGAGAAACCCAGACTGCATCTATACACAGCGGACAGAGAA---	623
QY	83	SerCysLeuArgGluGlyGlyLysGlyLysGlnAsnGlyAspSerSerAlaGlyLysAsp	102
Db	623	-----	623
QY	103	PheProProProAlaGluValGluProThrProGluAlaGluLeuAlaGluAlaProCys	122
Db	624	-----CGGCTGCCAGTACTGCCGTACACAGAGTCCCTAACCTCGCG-----	665
QY	123	HisAspSerGluAlaSerLysLeuGlyAla-----ProAlaAlaGlyGlyGlu	138
Db	666	CATGACGCCCAACAGCGTCCAGAGAGAGCGTCAACGCCGCCCGCAATCGCGCGCG---	722
QY	139	GluGluTrpGlyGlnGlnArg-----	146
Db	723	-----TAGGCTCAGCGCCACCGGAGCGGAGCGGAGTACCGCTCAGGTTCCGTAGCGCGGATC	776
QY	147	GlnLeuGlyLysLysLysHisArgArgArgProSerLysLysArgHisTrpLysPro	166
Db	777	CAGCTCTCAAGGCGGAGGAGGAGGAGCGG-----CGTTTCTGGCGGAT	821
QY	167	TyrTyrLysLeuThrTrpGluGlyLysLysLysPhe--AspGluLysGlnSerLeuArgAl	186
Db	822	-----GGGCACGCGCAACGCTTCTATGACTACTTCATGACCAATAGGCT	863
QY	186	AspArg-----IleArgAlaGluMetPheAlaLysGlyGln--	198
Db	864	GTTCCAGGGAATTTCTGATCTGAGCGCATCATAGAGCGGAGCAGGAGGAGAACCAATG	923
QY	199	-----ProValAlaProTyrAsnThrTrpGlnPheLeuMe	210
Db	924	CGGCGATTCGTCAGTACGATTCTCGCGCGTTGGTTCCTATTTCACACAGTCCAG-----	975
QY	210	TAspAspHisAspGlnGluIleuProAspLysLysThrLysLeu-----	224
Db	976	-----CCGGACTACAAGGATCCGCTGTCCGCGCTGTGCCAAGT	1013
QY	225	-----TyrSerLysArgAlaAlaAlaLysSerAs	234
Db	1014	GGTCAACAACAGCTCTTCAGATGTGTCGATACGCGCGCATGATGGCGCACTTGGCCA	1073
QY	234	pasPThSerAspAspAspPhe-----	241
Db	1074	GGTCCCGCTGGACAGCAGGTGATTCTGCTGAAGCCGTTGATCGAGCTGTCTATTGC	1133
QY	242	-----MetGluGluGlyGlyGluGluAspGlyGlySer	252
Db	1134	GAACGTGCGCTGGTGCACATCGTTTGGCTGGATGACGCGCGT---GCCGCGCGCGGCG	1190

```

Oy      252  TAspGLyMeGSLyGLyAspGLySer----- 260
           |||::||| ||||| |||
Db      1191 CGGTGGACATGAGCCACACATGGCTCTTTGACACGACATCACCGGGCTTCAGCCCCAGCA 1250
           |||::||| ||||| |||
Oy      261  -GLuPheLengLlaArgAspPheSerGLuThrTyrgLlaArgTyrfHisThrGLuSerLleu-- 279
           ||||| ::| ||||| |||||
Db      1251 GCCTTCTCTCAACACAGAGCTTCTGC-----TACATCGCAACAGATGGAT 1295
           ||||| ||||| |||
Oy      279  ----- 279
Db      1296 CAAGCCGGTGTGTCAACCATCTTCGACCGCATATGTGCGAGCTGATGTAAAGATGAA 1355
Oy      280  ----GLAspMetSerLysGlnGLuLau----- 287
           |||::| ::||| |||||
Db      1356 GCGGCTGAATCTTCACCGACGCGAGCTGTCTGCTTAAGGCCATCTACTGTACAAACC 1415
           |||::| ::||| |||||
Oy      288  -----LleLysGLuTyrgLengLlueGluLueGLyLysCysLueSerArgMetGLu-- 302
           ||||| ||||| |||
Db      1416 GGACATATCGCGGGATCAAGAGCGGGCGGAGATGAGATGCGCGGAGAAAGGTGTACGC 1475
           ||||| ::||| |||
Oy      303  -----AspGLuAsnAsnArgLueAArgLueGLuSerLysArgLengLylAspAspAl 320
           |||||::| |||||
Db      1476 TTGGCTGTGAGAGACATCGCCGCTGGAAACATCCG-----GGCGACGATGG 1520
           |||||::| |||||
Oy      320  aArgValaArgGLuLengLlueGluLueAspArg-LleArgAlaGLuAsnLengLlueL 340
           |||::||| ||| ||| |||||::|
Db      1521 ACGGTTTGGCACTGCTGCTGCTGCGCGCTTGCGCATGATCAGCTTGAGATGCC 1580
           |||||::| |||||
Oy      340  euThrGLuAsnGLuLueHisArgGLuGlnGLuAArgAlaProLueSerLys 356
           ||| ||| |||||
Db      1581 AGATGATCACTGTCTCTGCTCCGATTTGACAGCGACCGCGGCTGGAGAG 1630
           ||| ||| |||||

RESULT 8
US-08-464-514-1
: Sequence 1, Application US/08464514
: Patent No. 6265173
: GENERAL INFORMATION:
: APPLICANT: EVANS, RONALD M.
: APPLICANT: MCKEOWN, MICHAEL B.
: APPLICANT: ORO, ANTHONY E.
: APPLICANT: SEGRAVES, WILLIAM A.
: APPLICANT: YAO, TSO-PANG
: TITLE OF INVENTION: MULTIMERIC FORMS OF MEMBERS OF THE
: TITLE OF INVENTION: STEROID/THYROID SUPERFAMILY OF RECEPTORS WITH THE
: TITLE OF INVENTION: ULTRASPIRACLE RECEPTOR
: NUMBER OF SEQUENCES: 29
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PRETTY, SCHROEDER, BRUEGEMANN & CLARK
: STREET: 444 South Flower Street, Suite 2000
: CITY: Los Angeles
: STATE: California
: COUNTRY: United States
: ZIP: 90071
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/464,514
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/907,908
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Reiter, Stephen E.
: REGISTRATION NUMBER: 31192
: REFERENCE/DOCKET NUMBER: P41 9321
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 546-4737
: TELEFAX: (619) 546-9392

```



FILING DATE: 435  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/907,908  
 FILING DATE: 02-JUL-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Reiter, Stephen E.  
 REGISTRATION NUMBER: 31192  
 REFERENCE/DOCKET NUMBER: P41 9321  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (619) 546-4737  
 TELEFAX: (619) 546-9392  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2304 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 163..1701  
 US-08-486-403-1

Alignment Scores:  
 Pred. No.: 0.00669 Length: 2304  
 Score: 141.00 Matches: 106  
 Percent Similarity: 32.43% Conservative: 49  
 Best Local Similarity: 22.18% Mismatches: 127  
 Query Match: 7.38% Indels: 197  
 Gaps: 22

US-09-972-758a-2 (1-359) x US-08-486-403-1 (1-2304)

QY 13 ProGlnThrSerAsnCysThrGlyAlaAlaAlaValGlnGlnGlnLeuAsnProGlnArg 32  
 DB 388 CCCAATTGGCTGGAGGCTCTGCCGCCCTGCTCACTCCAGCCAGCATATCCGCTTACCATT 447  
 QY 33 ProProGlnAlaGlnGlnArg-----ValProGlnGlnAspSerArgTrpGlnSer 49  
 DB 448 CCGCTGAGCGGCGAGCAGCACCTCTGCTCTATTGG-CGGGATGGGGCGAGTGGCAACCA 506  
 QY 50 ArgAlaPheProGlnLeuGlyGlyArgProGly-----ProGlnGlyGlnGly 65  
 DB 507 CTACGGCGCTGACAGCTGTGAGGGCTGCAGAGGCTTCTTAAACGCACAGTGCACCAAGA 566  
 QY 66 Ser-----LeuGlnSerGlnProProProLeuGlnThrGlnAlaCysProGlnSer 82  
 DB 567 TCTCACATACGCTTGCAGAGGAGAACCGCACTGCATCTAGACAAAGCGGAGAGGAA--- 623  
 QY 83 SerCysLeuArgGlnGlnGlyGlnGlyGlnAsnGlyAspSerSerAlaGlyGlyAsp 102  
 DB 623 ----- 623  
 QY 103 PheProProAlaGlnValGlnProThrProGlnAlaGlnLeuLeuAlaGlnProCys 122  
 DB 624 -----CCGCTGACAGTACTCCGCTACCAAGAGTCCCTAACCTGCGG--- 665  
 QY 123 HisAspSerGlnAlaSerLysLeuGlyAla-----ProAlaAlaGlnGlyGln 138  
 DB 666 CATACACCGCAGACCGTCCAGAGGAGACGCTCAACAGCGGCGCCCGCAATGCGGCGG--- 722  
 QY 139 GlnGlnTrpGlnGlnGlnArg----- 146  
 DB 723 -----TAGGCTCAGCGCAGCAGGAGGCGCAGTACGCGTCCAGTTCGTAAGCGGATC 776  
 QY 147 GlnLeuGlnLysLysLysHisArgArgProSerLysLysLysArgHisTrpLysPro 166  
 DB 777 CAGCTTCACAGCGGAGAGAGAGGAGCGG-----CGTTCTGCGCGGAAT 821  
 QY 167 TyrTrpLysLeuThrTrpGlnGlnLysLysPhe-AspGlnLysGlnSerLeuArgAl 186  
 DB 822 -----GGCAGCGGCAACGGTTCTGATGACTTTCATGACCAATAGCGT 863

QY 186 aserArg-----IleArgAlaGlnMetPheAlaLysGlyGln-- 198  
 DB 864 GTCCAGGATTTCTTCGATCGAGCGCATCATAGAGCCGACAGCAGCGGAGACCAATG 923  
 QY 199 -----ProValAlaProTyrAsnThrGlnPheLeu 210  
 DB 924 CGGCGATGCTGACTGACCTGCTCCGCGTGTGGTCCATTATCCACAGTCAG----- 975  
 QY 210 LaspAspHisAspGlnGlnGlnProAspLeuLysThrGlyLeu----- 224  
 DB 976 -----CCGACTACAAAGGTCCTGTGCGGCTGTGCCAAGT 1013  
 QY 225 -----TyrSerLysArgAlaAlaAlaLysSerAs 234  
 DB 1014 GGTCAACAACAGCTTTCAGATGTCGATACCGCGGCAATGCCGCACTTGGCCA 1073  
 QY 234 PaspThrSerAspAspPhe----- 241  
 DB 1074 GGTGCGCTGGACACAGTGATTTGCTGAAGCCGCTTGATCGAGCTGCTCATTTGC 1133  
 QY 242 -----MetGlnGlnGlyGlnGlnAspGlyGlyse 252  
 DB 1134 GAACGTGGCTGTGTCAGCATCGTTTGGTGTGATGACCGCGT---GCCGCGCGGCGG 1190  
 QY 252 rAspGlyMetGlyLysGlySer----- 260  
 DB 1191 CGGTGACATACGCCAGCATGCTCTTTCAGCAGCATCACCAGGCGCTTCAGCCCAAGA 1250  
 QY 261 -GlnPheLeuGlnArgAspPheSerGlnTrpGlnArgTyrHisThrGlnSerLeu-- 279  
 DB 1251 GCTGTCTCTCAACAGAGCTTCTCG-----TACCATGCAACAGCTCGCAT 1295  
 QY 279 ----- 279  
 DB 1296 CAAGCCGCTGTACGCCATCTTCGACCCGATTTGTCGAGCTGATGATAAGATGAA 1335  
 QY 280 -----GlnAsnMetSerLysGlnGlnLeu----- 287  
 DB 1356 GCGGCTGAATCTGCACGAGCGAGAGCTGTCTGCTGAAGCCATCATCTGTCACACCC 1415  
 QY 288 -----IleLysGlnTrpLeuGlnGlnGlyLysCysLeuSerArgMetGln-- 302  
 DB 1416 GGACATACCGCGGATCAAGAGCCGCGGAGATGAGATGCGCCGCAAGAGGTGTACGC 1475  
 QY 303 -----AspGlnAsnAsnArgLeuArgLeuGlnSerLysArgLeuGlnGlyAspAspAl 320  
 DB 1476 TTGCTGAGCAGCAGCTGCGCTGGAACATCCG-----GGCGACGATGG 1520  
 QY 320 ArgValArgGlnLeuGlnGlnLeuAspArg-LeuArgAlaGlnAsnLeuGlnLeu 340  
 DB 1521 ACCTTTGCCCACTGCTGCTGCTGCTGCGCGCTTGGCATCATGAGCTGAAGTGC 1580  
 QY 340 eutThrGlnAsnGlnLeuHisArgGlnGlnGlnAlaArgProLeuSerLys 356  
 DB 1581 AGGATCACTCTTCTCTTCCGATTTACACGAGCAGCGGCGCTGAGAGAG 1630

RESULT 10  
 US-09-265-013-2  
 Sequence 2, Application US/09265013  
 Patent No. 6451304  
 GENERAL INFORMATION:  
 APPLICANT: FRIEDMAN, Theodore  
 APPLICANT: MIYANOHARA, Atsushi  
 TITLE OF INVENTION: METHOD FOR RETROVIRUS VECTOR PRODUCTION BY SEPARATED  
 TITLE OF INVENTION: GAG AND POL EXPRESSION  
 FILE REFERENCE: 041673/2010  
 CURRENT APPLICATION NUMBER: US/09/265,013  
 NUMBER OF SEQ ID NOS: 5  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 2  
 LENGTH: 1617



```

: TYPE: DNA
: ORGANISM: Moloney murine leukemia virus (MOLV)
: US-09-265-013-2

Alignment Scores:
  Pred. No.:      0.00899      Length:      1617
  Score:          137.00      Matches:      90
  Percent Similarity: 30.63%   Conservative: 42
  Best Local Similarity: 20.88%  Mismatches: 152
  Query Match:      7.17%     Indels:      147
  DB:                4        Gaps:      17

US-09-972-758a-2 (1-359) x US-09-265-013-2 (1-1617)

OY      18 CysThrGly---AlaIaIaIaValGInGluGluLeuAsnProGluArpProGluAl 36
      |||||  |||:|||||
Db      300 TGTACACCTTAAGCTTCGCCCTCTCTCTCCATCGCGCGCTCTCTCCCTTGAAAC 359
OY      36 aGluGluArGValProGluGluAspSerArGTrpGInSerArGAlaPheProGluLeuG 56
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      360 TCTCGTTCGACCCCGCTCGATCCTCCCTTATTCAGCCCTCAGT---CCTTCTAGG 416
OY      56 yGluArGProGluProGlu-----GlyGluGlySer-----LeuG 68
      |  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      417 CGCAAAACCTAAACCTCAAGTCTTCTGACAGTGGGGCGCCGTCATCGACCTACTTAC 476
OY      68 userGInProProProLeuGInThrGInAlaCysProGluSerSerCysLeuArGLeuG 88
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      477 AGAAGACCCCCCGCTTATAGGAGCCCAAGACCACCCCTTCGAC-----AGGAGCG 530
OY      88 yGluGlyGlyGInAsnGlyAspAspSerSerAlaGlyGlyAspPheProProProAlaG 108
      |  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      531 A-----AATGTCGAGAGAGAGACCCCTGGGGGAGAGACGCG----- 567
OY      108 uValGluProThProGluAlaGluLeu-----AlaGInProCysHisAs 124
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      568 ----GACCCCTCCCAATGACATCTCCGCTACCTGGAGAGAGAGACCCCTTGCGCA 623
OY      124 pSerGluAlaSerLeuGlyLeuGlyAlaProAlaGInGlyGluGluGluTrpGluGInG 144
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      624 CTCACACTACCTCGCAGGATTCGCCCTCGCGGAGAGAGAAC----- 666
OY      144 nGluArGInLeuGlyLysLysLysHisArGArGArProSerLysLysLysArGHisTr 164
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      667 -----GGACAGCTTCATATGCGCGCTTCTCTCTGACACTTACAACTG 713
OY      164 pLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGluLysGInSerLe 184
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      714 GAAA-----AATATAACCTCTCTTTTCTGAGATCCAGGTAA 752
OY      184 uArGAlaSerArGInleArGAlaGluMetPheAlaLysGlyGInProValAlaProTyrAs 204
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      753 ACTGACAGCTGTATGACGTCTTCTCATC----- 783
OY      204 nThrTrGInPheLeuMetAspAspHisAspGInGluProAspLeuLysThrGly-- 223
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      784 -ACCATCAAGCCACTGGGAGACATGTGACGACGTGGGAGCTCTGTCGACCGAGAG 842
OY      224 -----LeuTyrSerLysArGAlaIaIaLysSerAspAsp----- 235
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      843 AGAAAAACAACGGCTCTCTTAAGGCTAGAAAGGCGTCCGGGGGATGATGAGCGGCC 902
OY      236 -----ThSerAspAspAsp 241
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      903 CACTCAACTGCCCAATGAAGTCATCCGCTTTTCCCTTCGAGCGCCAGACTGGAGATTAA 962
OY      241 eMetGluGluGlyGluGlu----- 248
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      963 CACCAACCAAGGAGGTAGGAACACACTAGTCACATTCGACAGTTCCTCTACGGGTCT 1022
OY      249 -----AspGlyGlySerAspGlyme 255
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      1023 CCAAAAGCGCGGAGAGGCCCAACAATTGGCCAGGTAAAGAAATACACAGCGGCC 1082

```

```

OY      255 tGlyGlyAspGlySerGluPheLeuGInArGAspPheSerGluThr-TyrGluArGTr-- 274
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      1083 CAATGAGTCTCCTCGGCCCTTCTTAGAGAGA---CTTAAGAGAGCTATCGACGTACAC 1139
OY      274 ----- 274
Db      1140 TCTTATGACCCCTGAGGACCCAGAGCAGAAACTAATGTGTATGTCTTCTTATTTGGCA 1199
OY      275 -----HisThrGluSerLeuGInAsnMetSerLy 284
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      1200 GTCTGCCCCAGACATTGGAGAAAGTTAGACAGCTTAGAAGATTAAAAACAGACCT 1259
OY      284 sGInGluLeuLeuLeuGlyTrpLeuGluGluGluGluGluGluGluGluGluGluGlu 304
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      1260 TGGAGATTGTGGTTAGAGAGCAGAAAGATCTTTAATTAACGAGAAACCCCGAAGAAAG 1319
OY      304 uAsnAsnArGLeuArGLeuGluSerLysArGLeuGlyGlyAspAspAlaArGValArG 324
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      1320 ACAGGAACGTATCAGC-----AGAGA 1340
OY      324 uLeuGluLeuGluLeuAspArGLeuArGAlaGluAsnLeuGInLeuLeuThrGluGInG 344
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      1341 AACGAGAGAAAAAGAGAACGCCGTAGACAGAGATGACAGAAAGAGAAAGAGAGA 1400
OY      344 uLeuHisArGInGInGluGluArGAlaProLeu 354
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      1401 TCGTAGGAGCATATGAGAGATGAGCAAGCTA 1431

RESULT 11
: US-09-11-745-3
: Sequence 3, Application US/09011745
: Patent No. 6165715
: GENERAL INFORMATION:
: APPLICANT: Collins, Mary KT
: APPLICANT: Weis, Robin A
: APPLICANT: Takeuchi, Yasuhiro
: APPLICANT: Cosset, Francois-Loic
: TITLE OF INVENTION: Expression systems
: FILE REFERENCE: 09/011,745
: CURRENT APPLICATION NUMBER: US/09/011,745
: EARLIER FILING DATE: 1998-06-22
: EARLIER APPLICATION NUMBER: PCT/GB96/02061
: EARLIER FILING DATE: 1996-08-23
: EARLIER APPLICATION NUMBER: GB9517263.1
: NUMBER OF SEQ ID NOS: 29
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 3
: LENGTH: 7308
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Portion of
: OTHER INFORMATION: construct
: US-09-011-745-3

Alignment Scores:
  Pred. No.:      0.0723      Length:      7308
  Score:          137.00      Matches:      90
  Percent Similarity: 30.63%   Conservative: 42
  Best Local Similarity: 20.88%  Mismatches: 152
  Query Match:      7.17%     Indels:      147
  DB:                4        Gaps:      17

US-09-972-758a-2 (1-359) x US-09-011-745-3 (1-7308)

OY      18 CysThrGly---AlaIaIaIaValGInGluGluLeuAsnProGluArpProGluAl 36
      |||||  |||:|||||
Db      1866 TGTACACCTTAAGCTTCGCCCTCTCTCTCCATCGCGCGCTCTCTCCCTTGAAAC 1925
OY      36 aGluGluArGValProGluGluAspSerArGTrpGInSerArGAlaPheProGluLeuG 56
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

Db 1926 TCCTGCTGACCCCGCTGCATCTCCCTTATTCAGCCCTCACT---CCTTCTTAGG 1982
QY 56 YGATGPPGPGIYProGlu-----GlyGluGlySer-----LeuG1 68
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 1983 CGCCAAACCTAAACCTCACTTCTTCTGACAGTGGGGGCGCTCATGCACCTACTTAC 2042
QY 68 uSerGlnProProLeuGlnThrGlnAlaCysProGluSerSerCysLeuArgGluG1 88
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2043 AGAAGACCCCGCTTATAGGACCAAGACACCCCTTCCGAC-----AGGACGG 2096
QY 88 YGluLYGlyGlnAsnGlyAspAspSerSerAlaGlyGlyAspPheProProAlaG1 108
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2097 A-----AATGTGGAGAGAGCCCTCGGGAGAGCACCG----- 2133
QY 108 uValGluProThrProGluAlaGluLeuLeu-----AlaGlnProCysHis 124
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2134 ----GACCCCTCCCAATGCATCTCGCTACAGTGGAGACGGAGACCCCTGTGGCGCA 2189
QY 124 pSerGluAlaSerLeuGlyLeuGlyAlaProAlaAlaGlyGlyGluGluGluTrpGlyGlnG1 144
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2190 CTCCTACTACTCCGACGATTCCTCCCTCCGCGAGAGCAAC----- 2232
QY 144 ngInArGlnLeuGlyLysLysLysHisArgArgArgProSerLysLysArgHisTr 164
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2233 -----GGACAGCTTCAATACGCGCTTCTCTCTTCTGACCTTACACTG 2279
QY 164 pLysProTYrTYrLysLeuThrTrpGluGluLysLysPheAspGluLysGlnSerLe 184
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2280 GAAA-----AATATACCTCTCTTTTCTGACAACTCAGGTAA 2318
QY 184 uAlaGlnAlaSerAlaIleArgAlaGluMetPheAlaLysGlyGlnProValAlaProTYrAs 204
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2319 ACTGACAGCTGTGATGAGTGTCTCTCATC----- 2349
QY 204 nThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeuLysThrGly-- 223
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2350 -ACCCATGACCCCACTGGAGACACTGTCTACAGCTGTGGAGCTCTGCTGACCGGAGA 2408
QY 224 -----LeuTYrSerLysArgAlaAlaAlaLysSerAspSph----- 235
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2409 AGAAACAAACAGGGGTGCTTACAGGCTAGAAAGCGGTCGCGGGGATGATGCGGCC 2468
QY 236 -----ThSerAspAspSph 241
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2469 CACTCACTCCCAATGAATGTCATGCCGCTTTCCCTCGACGCGCCACACATCGGATTA 2528
QY 241 eMetGluGluGlyGluGlu----- 248
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2529 CACACACCCAGAGGTAGGACCACTAGTCACATCGCACTTGTCTTACGGGTCT 2588
QY 249 -----AspGlyGlySerAspGlyMe 255
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2589 CCAAACGGGGGAGAGAGCCCAATTTGGCCAAAGGTAAAGGATACACAAAGGCC 2648
QY 255 tGlyGlyAspGlySerGluPheLeuGlnArgAspPheSerGluThrTYrGluArgTYr-- 274
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2649 CAATGAGTCTCCCTCGGCTTCTTCTAGAGAGA---CTTAAGAGAGCTATCGCAGTACAC 2705
QY 274 ----- 274
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2706 TCCTTATGACCTGAGGACCCAGGCAAGAACTAATGTCTATGTCTTCTTCAATTGGCA 2765
QY 275 -----HisThrGluSerLeuGlnAsnMetSerLys 284
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2766 GTCTGCCCCAGACATTGGAGAAAGTTAGAGAGTTAGAAAGATTAAACAAAGACGT 2825
QY 284 sGlnGluLeuIleLysGluTYrLeuGluLeuGluLysCysLeuSerArgMetLysG1 304
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2826 TGGAGATTGGTTAGAGAGCAGAAAGATCTTTAATAACGAGAAACCCCGAAGAAAG 2885
QY 304 uAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAlaArgAlaArgG1 324
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2886 ACAGGAACGATATCAG-----AGAGA 2906

```

```

QY 324 uLeuGluLeuGluAspArgLeuArgAlaGlnAsnLeuGlnLeuThrGlnAsnG1 344
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2907 AACAGACGAAAGAAAGAACAGCCGCTTACAGAGAGATGACGACGAAAGAGAGAGA 2966
QY 344 uLeuHisArgGlnGlnGluArgAlaProLeu 354
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2967 TCCTAGACAGCATAGACATGACGAAAGCTA 2997

RESULT 12
US-09-011-745-4
; Sequence 4, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Collins, Mary KL
; APPLICANT: Weiss, Robin A
; APPLICANT: Takeuchi, Yasuhiro
; APPLICANT: Cosset, Francois-Lolc
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; EARLIER FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 7308
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Portion of
; US-09-011-745-4

Alignment Scores:
Pred. No.: 0.0723 Length: 7308
Score: 137.00 Matches: 90
Percent Similarity: 30.63% Conservative: 42
Best Local Similarity: 20.88% Mismatches: 152
Query Match: 7.17% Indels: 147
Gaps: 17

US-09-972-758a-2 (1-359) x US-09-011-745-4 (1-7308)

QY 18 CysThrGly---AlaAlaAlaValGlnGluLeuAsnProGluArgProPro-GlyAl 36
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 1866 TGTACACCTTAAGCCCTCGCTCTCTCTCTCCATCGGCCGCTCTCCCTTGAACC 1925
QY 36 aGluGluArgValProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuG1 56
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 1926 TCCTGCTTGCACCCCGCTCGATCTCTCTTATTCACACCTTCACCT---CCTTCTTAGG 1982
QY 56 YGATGPPGPGIYProGlu-----GlyGluGlySer-----LeuG1 68
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 1983 CGCCAAACCTAAACCTCAAGTCTTCTGACAGTGGGGGCGCTCATGCACCTACTTAC 2042
QY 68 uSerGlnProProProLeuGlnThrGlnAlaCysProGluSerSerCysLeuArgGluG1 88
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2043 AGAAGACCCCGCTTATAGGACCAAGACACCCCTTCCGAC-----AGGACGG 2096
QY 88 YGluLYGlyGlnAsnGlyAspAspSerSerAlaGlyGlyAspPheProProAlaG1 108
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2097 A-----AATGTGGAGAGAGCCCTCGGGAGAGCACCG----- 2133
QY 108 uValGluProThrProGluAlaGluLeuLeu-----AlaGlnProCysHis 124
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2134 ----GACCCCTCCCAATGCATCTCGCTACAGTGGAGACGGAGACCCCTGTGGCGCA 2189
QY 124 pSerGluAlaSerLeuGlyLeuGlyAlaProAlaAlaGlyGlyGluGluGluTrpGlyGlnG1 144
| : : : : : | : : : : : | : : : : : | : : : : : |

```

```
Dh 2190 CTCACACTGCGAGGACATTCCTCCGCGAGGAGCAAC----- 2232
Qy 144 nglArglnLeuLysLysHisArgArgProSerLysLysArgHisTr 164
Db 2233 -----GGACAGCTTCAATACTAGCCGCTTCCTCTTGTGACTTTACACTG 2279
Qy 164 pLysProTyrTyrLysLeuThrTPGluGluLysLysPheAspGluLysGlnSerLe 184
Db 2280 GAAA-----AATAATACCCCTTCTTCTGTAAGATCCAGCTAA 2318
Qy 184 uArgAlaSerArgIleArgAlaGluMetPheAlaLysGluProValAlaProTyrAs 204
Db 2319 ACTGACACTCTGTGATCGAGTCTGTTCTCATC----- 2349
Qy 204 nThrThrGlnPheLeuMetAspAspHisAspGlnGluLupProAspLeuLysThrGly-- 223
Db 2350 -ACCATTCAGCCCACTGGGAGCATGTCACACACTGTGGGAGACTGTGCTGACCGGAGA 2408
Qy 224 -----LeuTyrSerLysArgAlaAlaLysSerAspSph----- 235
Db 2409 AGAAAAACACAGCGGTCTCTTAGAGCGTAGAAAGCGGTGCGGGCGCATGATGGGCGCCC 2468
Qy 236 -----ThrSerAspAspSph 241
Db 2469 CACTCACTGCCAATGAATGATGCCGCTTTTCCCTCGAGCGCCGACACTGGAGTTA 2528
Qy 241 eMetGluGluLysGluGlu----- 248
Db 2529 CACCACCCAGGAGAGGACCAACCACTAGTCCACTATGCCAGTTGCTCTTACGGGGTCT 2588
Qy 249 -----AspGluLysSerAspGlyMe 255
Db 2589 CCAAAACGCGGCGAGAACCCCAATTTGGCCAGTAAGTAAGTAATACACAAGGCC 2648
Qy 255 tGluGluAspLysSerGluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyr-- 274
Db 2649 CAATGAGTCTCCCTGGCTTCCTAGAGAGA---CTTAAGGAGCCATTCACAGTTACAC 2705
Qy 274 ----- 274
Db 2706 TCCATTATGACCTTAGAGGACCCAGGCAAGAACTAATCTGTCTATGTTCTTATTTGGCA 2765
Qy 275 -----HisThrGluSerLeuGlnAsnMetSerLys 284
Db 2766 GTCTGCCCCAGACATTGGAGAAAGTTAGACAGCTTAGACATTTAAATAACAAGACCT 2825
Qy 284 sGlnGluLeuLysGluTyrLeuGluLysCysLeuSerArgMetGluAspGlu 304
Db 2826 TGGAGATTGGTTAGAGAGGCAAGAAAGATCTTTAATAAACGAGAAACCCCGCAAGAAC 2885
Qy 304 uAsnAsnArgLeuArgLeuGlnSerLysArgLeuGluGlyAspAspAlaArgValArgG 324
Db 2886 AGAGAGAACGTATCAG-----ACAGA 2906
Qy 324 uLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeuThrGluAsnG 344
Db 2907 AACAGAGGAAACCAAGAACCCCTAGAGACAGAGATGACAGAAAGCAAGAAAGAGA 2966
Qy 344 uLeuHisArgGlnGlnArgAlaProLeu 354
Db 2967 TCGTAGGAGACATAGAGATGATGACCAAGCTA 2997
```

```
RESULT 13
US-09-011-745-2
: Sequence 2, Application US/09011745
: Patent No. 6165715
: GENERAL INFORMATION:
: APPLICANT: Collins, Mary KL
: APPLICANT: Weiss, Robin A
: APPLICANT: Takeuchi, Yasuhiro
: APPLICANT: Cosset, Francois-Lolc
: TITLE OF INVENTION: Expression systems
: FILE REFERENCE: 09/011,745
```

```
: CURRENT APPLICATION NUMBER: US/09/011,745
: CURRENT FILING DATE: 1998-06-22
: EARLIER APPLICATION NUMBER: PCT/GB96/02061
: EARLIER FILING DATE: 1996-08-23
: EARLIER APPLICATION NUMBER: GB9517263.1
: EARLIER FILING DATE: 1995-08-23
: NUMBER OF SEQ ID NOS: 29
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 2
: LENGTH: 7616
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Portion of
US-09-011-745-2
```

## Alignment Scores:

```
Pred. No.: 0.0766 Length: 7616
Score: 137.00 Matches: 90
Percent Similarity: 30.63% Conservative: 42
Best Local Similarity: 20.88 Mismatches: 152
Query Match: 7.17% Indels: 147
DB: 4 Gaps: 17
```

US-09-972-758a-2 (1-359) x US-09-011-745-2 (1-7616)

```
Qy 18 CysThrGly---AlaAlaAlaValGlnGluLeuAsnProGluArgProGlyAl 36
Db 1027 TGTACACCTTAAGCTCCCTCCCTCTTCCATCCGCGCCCTCTCTTGAACC 1086
Qy 36 aGluGluArgValProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuG 56
Db 1087 TCTCTGTTGACCCCGCCCTGCATCTCTCTTATTCACCTCACT---CCTTCTTAGC 1143
Qy 56 yGluArgProGlyProGlu-----GlyGluGlySer-----LeuG 68
Db 1144 CGCCAACCTTAACCTCAAGTTCTTCTGACAGTGGGGGCGCTCATCGACCTACTTAC 1203
Qy 68 uSerGlnProProLeuGlnThrGlnAlaLysProGluSerSerCysLeuArgGluG 88
Db 1204 AGAAGACCCCGCCCTTATAGGAGCCAGACACCCCTTCCGAC-----AGGACGG 1257
Qy 88 yGluLysGlyGlnAsnGlyAspSerSerAlaGlyGlyAspPheProProAlaG 108
Db 1258 A-----AATGTTGAGAGAGGACCCCTGCGAGAGCACC----- 1294
Qy 108 uValGluProThrProGluAlaGluLeu-----AlaGlnProCysHisAs 124
Db 1295 -----GACCCCTCCCAATGCGATCTCGCTAGGAGAGGAGGACCCCTGTGGCGGA 1350
Qy 124 pSerGluAlaSerLysLeuGluAlaProAlaAlaGlyGlyGluGluLupTrpGlyGlnG 144
Db 1351 CTCACACTGCGAGGACATTCCTCCGCGGAGGAGAAAC----- 1393
Qy 144 nglArgGlnLeuLysLysHisArgArgProSerLysLysArgHisTr 164
Db 1394 -----GGACAGCTTCAATACTAGCCGCTTCTCTTCTGACCTTTACAACTG 1440
Qy 164 pLysProTyrTyrLysLeuThrTPGluGluLysLysPheAspGluLysGlnSerLe 184
Db 1441 GAAA-----AATAATAACCCCTTCTTCTGTAAGATCCAGCTAA 1475
Qy 184 uArgAlaSerArgIleArgAlaGluMetPheAlaLysGluProValAlaProTyrAs 204
Db 1480 ACTGACACTCTGTGATCGAGTCTGTTCTCATC----- 1510
Qy 204 nThrThrGlnPheLeuMetAspAspHisAspGlnGluLupProAspLeuLysThrGly-- 223
Db 1511 -ACCATTCAGCCCACTGGGAGCATGTCACACACTGTGGGAGACTGTGCTGACCGGAGA 1569
Qy 224 -----LeuTyrSerLysArgAlaAlaLysSerAspSph----- 235
```

```

Db 1570 AGAAAAACAAGGGGCTCTTAGAGGCTAGAAAGGCGGTGGGGGCGATGATGGGCCCC 1629
OY 236 -----ThrsAspAspAsp 241
Db 1630 CACTCACTGCCCAATGAATGCATGCCGCTTTTCCCTCGAGGCGCCAGACTGGGATTA 1689
OY 241 eMeTcGlnGlnGlyGlyGlnGlu----- 248
Db 1690 CACCACCCAGGAGAGGTAGGAACCACTAGTGCACATGTGCACAGTTGCTCCTAGCGGCT 1749
OY 249 -----AspGlySerAspGlyMe 255
Db 1750 CCAAAACGGGCGAGAAAGCCCAACCAATTTGGCCAGGTAAAGATTAACCAACAAGGCC 1809
OY 255 tGlyGlyAspGlySerGlnPheLeuGlnArgAspSerGlnThrTrpGlyuArgTyr-- 274
Db 1810 CAATGAGCTCTCCCTCGGCTTCTTAGAGAGA---CTTAGGAAGCTTATCCAGAGTACAC 1866
OY 274 ----- 274
Db 1867 TCCTTATGACCCCTGAGAGACCCAGGCAAGAAACTATGTCTATGTCTTCATTTGGCA 1926
OY 275 -----HisThrGlnSerLeuGlnAsnMetSerGly 284
Db 1927 GTCTGCCCCAGACATTGGGAGAAAGTTAGACAGTTTAAGAAATTTAAAAACAAGACCT 1986
OY 284 sGlnGlnLeuLeuLeuGlySerGlyTrpLeuGlnLeuGlnLysCysLeuSerArgMetGluAspG1 304
Db 1987 TGGAGATTGTGTAGAGAGCGCAGAAAGATCTTAAATTAACGAGAAACCCGGAGAAAG 2046
OY 304 uAsnAsnArgLeuArgLeuGlnSerLysArgLeuGlnGlyLysAspAlaArgValArgG1 324
Db 2047 AGAGGAACCTGTATCAGG-----AGAGA 2067
OY 324 uLeuGlnLeuGlnLeuAsnArgLeuArgAlaGluAsnLeuGlnLeuThrGlnAsnG1 344
Db 2068 AACAGAGAAAAAGAAAGAACGCCCGTAGGACACAGAGATGACGACAAAGAAAGAAAGAGA 2127
OY 344 uLeuHisArgGlnGlnGlnArgAlaProLeu 354
Db 2128 TCGTAGAGACATAGAGAGATGACCAAGCTA 2158

RESULT 14
US-08-258-420-13/c
; Sequence 13, Application US/08258420
; Patent No. 5710037
; GENERAL INFORMATION:
; APPLICANT: Nienhuis, Arthur W.
; APPLICANT: Vanin, Ello F.
; TITLE OF INVENTION: No. 5710037e1 Retroviral Envelope and LTR and Retroviral Vector
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gillfillan, Cecchi, Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/258,420
; FILING DATE: 10-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Olstein, Elliot M.
; REGISTRATION NUMBER: 24,025.

```

```

REFERENCE/DOCKET NUMBER: 271010-208
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 8202 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
FEATURE:
NAME/KEY: viral genome
US-08-258-420-13

Alignment Scores:
Pred. No.: 0.0848 Length: 8202
Score: 137.00 Matches: 90
Percent Similarity: 30.63% Conservative: 42
Best Local Similarity: 20.88% Mismatches: 152
Query Match: 7.17% Indels: 147
DB: 1 Gaps: 17

US-09-972-758A-2 (1-359) x US-08-258-420-13 (1-8202)
QY 18 CysThrPrGly--AlaAlaAlaValGlnGluGluAsnProGluArgProPro-GlyAl 36
||||| ||||| ||||| |||||
DB 6847 TGTACACCCCTAAGCTCCGCCCTCCCTCTTCCTCCATCCGCCGCCCGTCTCCCTTGAAAC 6788
36 aGluGluArgValProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuG 56
||||| ||||| ||||| |||||
DB 6787 TCCCTGGTTCGACCCCGCCCTCCGCTCCCTCCCTTATCCACCCCTCACT--CCCTCTCTAGG 6731
56 yGlyArgProGluArgProGlu-----GlyGluGlySer-----LeuG 68
| :||| ||||| |||||
DB 6730 CCCCACACTAAACCTCAAGTCTTTCTTGACAGTGGGGGCCCGCTCATCGACACTACTAC 6671
68 uSerGlnProProLeuGlnPheGlnAlaCysProGluSerSerCysLeuArgGluG 88
||||| ||||| ||||| |||||
DB 6670 AGAAGACCCCCCGCTTATAGGAGCCCAAGACCAACCCCTCCGCCAC-----AGGAGCG 6617
88 yGluLysGlyGlnAsnGlyAspAspSerSerAlaGlyLysPhePheProProAlaG 108
||||| ||||| ||||| |||||
DB 6616 A-----AATGGTGGAGAGGACACCCCTGGGAGAGGCACCG----- 6580
108 uValGluProThrProGluAlaGluLeuLeu-----AlaGlnProCysHisAs 124
:::||||| ||||| |||||
DB 6579 ---CACCCCTCCCAATGGCATCGCATCGCGCTACGTGGAGAGGGAGCCCGCTGTGGCGGA 6524
124 pSerGlnAlaSerLysLeuGluAlaProAlaAlaGlyLysGluGluGluTrpGlnG 144
||||| ||||| ||||| |||||
DB 6523 CTCGACCTACCTCGCGGCGCATTCGCCCTCCGCGGAGGAGGACAC----- 6481
144 ngLnaArgGlnLeuGlnGlyLysLysHisArgArgProSerLysLysArgHisTr 164
||||| ||||| ||||| |||||
DB 6480 -----GGACACCTTCATATAGTGGCCGCTTCCTCTCTGTACACTG 6434
164 pLysPheTrpGlyTyrGlyLysLeuPheTrpGluGlnLysLysPheAspGlnLysGlnSerLe 184
||||| ||||| ||||| |||||
DB 6433 GAAA-----AATAATAAACCTCTTTCTGTGAAGATCCAGGTAA 6395
184 uArgAlaSerArgTrpLeaArgAlaGluMerPheAlaLysGluLysProValAlaProTyrAs 204
::: ||||| ||||| |||||
DB 6394 ACTGACACCTGTGATGACGATGCTGTTCATC----- 6364
204 nThrThrGlnPheLeuHisCAspRpnHisAspGlnGluLysProAspLeuLysThrGly-- 223
||||| ||||| ||||| |||||
DB 6363 -ACCCATGAGGCCAOCSTGGAGACACSTGTCACACACSTGTGGGAGACTGCTGTACCCGAGA 6305
224 -----LeuTyrSerLysArgAlaAlaLysSerAspAsp----- 235
||||| ||||| ||||| |||||
DB 6304 AGAAACACACCGGTGCTCTTGAAGAGCTAGAAAGAGCGCTCGGAGGAGATGAGGAGGCC 6245

```

QY 236 -----ThrsrAspaspaph 241  
Db 6244 CACTCACTGCCCATGATGCATGCGCTTTCCCTCGAGGCCAGACTGGATTA 6185  
QY 241 eMeGluGluGlyGlyGluGlu----- 248  
Db 6184 CACCACCCAGCAGTAGGAACACACCTAGTCCACTATCGCAGTTGCTCTAGCGGGTCT 6125  
QY 249 -----AspGlySerAspGlyme 255  
Db 6124 CCAAAACCGCGGAGAACCCACCACCAATTGGCCAGCTAAAGAAATATACACAGGCGC 6065  
QY 255 LgLyGlyAspGlySerGluPheLeuGlnArgAspPheSerGluThrTyrgluArgTy-- 274  
Db 6064 CAATGAGTCTCCCTCGGCTCTCTAGAGAGA---CTTAAGAACCCCTATCGCAGCTTACAC 6008  
QY 274 ----- 274  
Db 6007 TCCTTATGACCCTGAGGACCCAGGCAAAACTAATGTGTATGTCTTTTCATTTGGCA 5948  
QY 275 -----HisthGluSerLeuGlnAsnMetSerly 284  
Db 5947 GTCTGCCCCAGACATTGGCAGAAAGTTAGAGAGGTTGAAACATTTAAACAAAGACGCT 5888  
QY 284 sGlnGluLeuLLeuGlyLeuGluLeuGluLysCysLeuSerArgMetGluAspG1 304  
Db 5887 TCGAGATTGTGTTAGACAGCGCAAAAGATCTTAATAACGACAAACCCCGGAGAAAG 5828  
QY 304 uAsnAsnArgLeuArgLeuGlnSerLysArgLeuGlyGlyAspAspAlaArgValArgG1 324  
Db 5827 AGAGGAACGTATACAG-----AGAGA 5807  
QY 324 uLeuGlnLeuGlnLeuAsnArgLeuArgAlaGluAsnLeuGlnLeuThrGluAsnG1 344  
Db 5806 AACGAGAGAAAACAAAGACCCCTAGCAGACAGAGATGACGAGAAGAAAGAAAGAGA 5747  
QY 344 uLeuHisArgGlnGlnGluArgAlaProLeu 354  
Db 5746 TCGTAGGAGACATGAGAGATGACGCAAGCTA 5716

RESULT 15  
US-08-850-961-1  
Sequence 1, Application US/08850961  
Patent No. 6013517  
GENERAL INFORMATION:  
APPLICANT: Respass, James G.  
APPLICANT: De Polo, Nicholas J.  
APPLICANT: Chada, Sunil  
APPLICANT: Sauter, Sybille  
APPLICANT: Bodner, Mordechai  
APPLICANT: Driver, David A.  
TITLE OF INVENTION: CROSSLESS RETROVIRAL VECTORS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Chiron Corporation, Intellectual Property - R440  
STREET: P.O. Box 8097  
CITY: Emeryville  
STATE: California  
COUNTRY: USA  
ZIP: 94662-8097  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/850,961  
FILING DATE: 05-MAY-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kruse, No. 6013517man J.  
REGISTRATION NUMBER: 35,235  
REFERENCE/DOCKET NUMBER: 930049, 424C4 / 1147,005

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-3520  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8332 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-850-961-1  
Alignment Scores:  
Pred. No.: 0.0867 Length: 8332  
Score: 137.00 Matches: 90  
Percent Similarity: 30.63% Conservative: 42  
Best Local Similarity: 20.88 Mismatches: 152  
Query Match: 7.17% Indels: 147  
Gaps: 17  
US-09-972-758A-2 (1-359) x US-08-850-961-1 (1-8332)  
QY 18 CysThrGly--AlaAlaAlaValGlnGluLeuAsnProGluArgProPro-GlyAl 36  
Db 920 TGTACACCCTAAGCCTCCGCTCTCTCCATCCGCGCCGCTCTCCCTTGAAC 979  
QY 36 aGluGluArgValProGluLysPheArgTrpGlnSerArgAlaPheProGlnLeuG1 56  
Db 980 TCTCTGTCGACCCCGCCTGCATCTCTCTTTATCCAGCCCTCACT---CTTCTCTAGG 1036  
QY 56 yGlyArgProGlyProGlu-----GlyGluGlySer-----LeuG1 68  
Db 1037 CGCCAAACCTAAACCTCAAGTCTTCTGTACAGTGGGGGCGCGCTCATCGACCTACTTAC 1095  
QY 68 uSerGlnProProProLeuGlnThrGlnAlaCysProGluSerSerCysLeuArgGluG1 88  
Db 1097 AGAAGACCCCGCGCTTATAGGACCCCAAGACCAACCCCTTCGCGAC-----AGGACGG 1150  
QY 88 yGluLysGlyGlnAsnGlnAsnSerSerArgAlaGlyAspPheProProAlaG1 108  
Db 1151 A-----AATGTGAGAGAGGACGCCCTCGCGGAGAGGACCG----- 1187  
QY 108 uValGluProThrProGluAlaGluLeuLeu-----AlaGlnProCysHisAs 124  
Db 1188 ---GACCCCTCCCATATGCGATCTGCTACGTGGAGACGGAGCGGCCCTGTGGCGCA 1243  
QY 124 pSerGlnAlaSerLysLeuGlnAlaProAlaAlaGlyGlyGluGluGluTrpGluGlnG1 144  
Db 1244 CTCACCTACCTCGAGGCAATTCCCGCTCCGCGAGGAGAAAC----- 1286  
QY 144 nGlnArgGlnLeuGlyLysLysLysHisArgArgProSerLysLysLysArgHisTr 164  
Db 1287 -----GGACAGCTTCAATAGCTGCGGCTTCTCTTCTGTACCTTTACACAGCTG 1333  
QY 164 pLysProTrpTyrlLysLeuThrTrpGluGluLysLysPheAspGluLysGlnSerLe 184  
Db 1334 GAAA-----AATATATACCCCTTTTCTTGTGAGATGCAGGTAA 1372  
QY 184 uArgAlaSerArgLLeuArgAlaGluMetPheAlaLysGlyGlnProValAlaProTyrrAs 204  
Db 1373 ACTGACAGCTCTGATCGACCTGTTCTTCATC----- 1403  
QY 204 nThrThrGlnPheLeuMetAspPheHisAspHisGlnGluProAspLeuLysThrGly-- 223  
Db 1404 -ACCATCAGCCCACTCGGAGCAGCTTCAGCAGCTTTTGGGAGCTGCTGTGACCGGAGA 1462  
QY 224 -----LeuTyrlSerLysArgAlaAlaAlaLysSerAspAsp----- 235  
Db 1463 AGAAACACAGGGGTCTTTAGAGCTTAGAAGGCGCTCGGCGCATGATGCGCGCC 1522  
QY 236 -----ThrsrAspaspaph 241  
Db 1523 CACTCACTGCCCATGATGATGCGCTTTTCCCTCGAGGCCAGACATGGGATTA 1582



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 10:19:50 ; Search time 80 Seconds  
(without alignments)  
924.636 Million cell updates/sec

Title: US-09-972-758a-2  
Perfect score: 1910  
Sequence: 1 MAEPFLSEYOHOPOTSNGT.....LTRENELHROERAPLSKPGD 359

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP\_TREMBL\_21:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhc:\*  
9: sp\_organelle:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp Vertebrate:\*  
14: sp Unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriapi:\*  
17: sp\_archaeapi:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1910	100.0	359	4	094992 homo sapien
2	1587.5	83.1	356	11	08409 mus musculu
3	1199	62.8	246	11	08VDF7 mus musculu
4	545.5	25.3	286	4	096MH2 mus musculu
5	483.5	25.3	280	11	09DAC7 mus musculu
6	209	10.9	349	5	09VPH2 mus musculu
7	150	7.9	1142	5	08T5B4 mus musculu
8	148.5	7.8	992	4	09NTH6 mus musculu
9	147	7.7	1455	4	09NTH6 mus musculu
10	142.5	7.5	560	6	095LS7 mus musculu
11	141.5	7.4	1192	5	09W475 mus musculu
12	140.5	7.4	443	5	076153 periplaneta
13	140	7.3	714	10	08TXA4 mus musculu
14	140	7.3	714	10	09ADE3 mus musculu
15	139	7.3	1520	4	015087 mus musculu
16	139	7.3	1781	4	09UKX0 mus musculu

17	139	7.3	1890	4	09UKW3 mus musculu
18	139	7.3	2073	4	09UKW2 mus musculu
19	138.5	7.3	944	11	092253 mus musculu
20	138.5	7.3	1665	11	08VT11 mus musculu
21	138	7.2	2072	4	08WYB5 mus musculu
22	137.5	7.2	540	11	09E053 mus musculu
23	136	7.1	399	10	094IK2 mus musculu
24	135.5	7.1	530	4	09UER6 mus musculu
25	135	7.1	540	4	096587 mus musculu
26	135	7.1	554	4	09HB02 mus musculu
27	135	7.1	602	3	09HEL8 mus musculu
28	133	7.0	462	11	09R1B7 mus musculu
29	133	7.0	538	15	09WJPD mus musculu
30	133	7.0	1144	5	09NJH7 mus musculu
31	133	7.0	1737	15	092808 mus musculu
32	132.5	6.9	530	4	073475 mus musculu
33	132.5	6.9	911	5	08T852 mus musculu
34	132	6.9	528	4	09H6U3 mus musculu
35	132	6.9	528	4	09H3P7 mus musculu
36	131	6.9	681	4	09P2N2 mus musculu
37	131	6.9	932	5	09VZP5 mus musculu
38	131	6.9	3484	5	P91257 mus musculu
39	130.5	6.8	628	11	008586 mus musculu
40	130.5	6.8	778	4	09UQ89 mus musculu
41	130.5	6.8	819	4	096ST2 mus musculu
42	129.5	6.8	1437	4	09NU94 mus musculu
43	129.5	6.8	1482	4	09NU93 mus musculu
44	129	6.8	1054	3	094097 mus musculu
45	129	6.8	1579	11	099WPI mus musculu

## ALIGNMENTS

RESULT 1	ID	094992	PRELIMINARY;	PRT;	359 AA.
AC	094992:				
DT	01-MAY-1999 (TREMBLrel. 10, Created)				
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)				
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)				
DE	HIS1 protein (HMBB-inducible).				
GN	HIS1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_taxid=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Kushnara M., Nagasaki K., Kimura K., Ishikawa S., Maass N., Ogawa S.,				
RA	Yamaguchi K.;				
RT	"Cloning of HMBB-inducible transcript, HIS1, in human vascular smooth				
RT	muscle cells.";				
RL	Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.				
RP	[2]				
RN	SEQUENCE FROM N.A.				
RC	TISSUE=LUNG;				
RA	Strausberg R.;				
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AB021179; BAA36166.1; -;				
EMBL	EMBL; BC006460; AAH06460.1; -;				
SO	SEQUENCE 359 AA: 40623 MW; B12845C4E2595FF0 CRC64;				
Query Match	100.0%;	Score 1910;	DB 4;	Length 359;	
Best Local Similarity	100.0%;	Pred. No. 1.7e-128;			
Matches 359;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
DB	1	MAEPFLSEYOHOPOTSNGTGAAYVDELNPERPPGAEEYVPPDSWOSRAPOGGRG 60			
QY	1	MAEPFLSEYOHOPOTSNGTGAAYVDELNPERPPGAEEYVPPDSWOSRAPOGGRG 60			
QY	61	PEEGSLSEOPPLQTOACPESSCLREGKGGGDDSSAGGDFPPAEVETPEALLQ 120			

D6	6	PEGEGLSEOPPLQACPESSCLREBKQNGDDSSAGGDPPPAVEPTPEALLAQ	120
QY	121	PCHDSSASKLGAAPAAEGEEMQOOROLGKKKHHRRPSKKRRHKPKYKLTWEKKKFE	180
D6	121	PCHDSSASKLGAAPAAEGEEMQOOROLGKKKHHRRPSKKRRHKPKYKLTWEKKKFE	180
QY	181	KSLRSRSRLRAEMFAKGPVAPYNTTQFLMDHDDEPDLTGLYSKRAAKSDOTSDDD	240
D6	181	KSLRSRSRLRAEMFAKGPVAPYNTTQFLMDHDDEPDLTGLYSKRAAKSDOTSDDD	240
QY	241	FMEEGGEEDGSDGMGDSFEQLQDFSETYERHTESLQMSKOEILKEYLELEKLSR	300
D6	241	FMEEGGEEDGSDGMGDSFEQLQDFSETYERHTESLQMSKOEILKEYLELEKLSR	300
QY	301	MEENNRLLPLESKRRLGGDDARVRELLELDRLRAENQLLTENELHQOENAPLSKSGD	359
D6	301	MEENNRLLPLESKRRLGGDDARVRELLELDRLRAENQLLTENELHQOENAPLSKSGD	359

RESULT	2
06R409	
ID	06R409
AC	06R409;
DT	01-JUN-2002 (Tremblrel. 21, Created)
DT	01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE	Cardiac lineage protein 1.
GN	Clp1.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxID=10090;
RN	[1]
RC	SEQUENCE FROM N.A.
RC	STRAIN=129/SVU;
RA	Huang F., Wagner M., Siddiqui M.;
RT	"Structure, expression, and functional characterization of the mouse
RT	Clp-1 gene.";
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RL	EMBL: AY090614; AAC09026.1.
SO	SEQUENCE 356 AA; 40243 MW; 242DEE66BA293B CRC64;

Query Match	83.1%;	Score 1587.5;	DB 11;	Length 356;
Best Local Similarity	-85.8%;	Pred. No. 1.7e-105;		
Matches 308; Conservative	11;	Mismatches 37;	Indels 3;	Gaps 2;

QY	I	MAPPFLSEVOHONOTSNCTGAAAVOELNPERPRAEEVYPREDSWOSRATPOLGSRPG	60
		:       :       :       :       :	
Db	1	MAPPPLTLEHHOHOTSNCSTGAAYVHEEHTSERPPSAEEVPKEDSWOSRASLOSGRPG	60
		:       :       :       :       :	
QY	61	PCEGGSLSESPPLLOTACPESSCLKEGKGOMDDSGSGRPPEAUEPTPEALLAO	12
		:       :       :       :       :	
Db	61	QESEGGLKHOLLPLLOTNACPETLSLKEGKGQGEDLTSG--ASPAEEPPSES--LVQ	11
		:       :       :       :       :	
QY	121	PCGDSEASKLGAPAAAGEEEMWGQQOLKKKKHRRRRSKKKHKWKYYVLTMEEKKFE	18
		:       :       :       :       :	
Db	118	PGHDSEATGOENARAAGGEPMWQQOLKKKKHRRRRSKKKHKWKYYVLTMEEKKFE	17
		:       :       :       :       :	
QY	181	KOSLRASRLRAEMFAKGPVARYNTQTLMDDHOEDPDLKTGLYSKRAAASDSTDSD	24
		:       :       :       :       :	
Db	178	KOSLRASRYAEKFAGGVAPARYNTQTLMDDHOEDPDLKTGLYPKRAAKSDSTDSD	23
		:       :       :       :       :	
QY	241	FMEEGEEDDGSGMGDGSSEFIQORFSETTEKYHTESIQNNNSKOFLIKEYELEKCLR	30
		:       :       :       :       :	
Db	238	FVEEAEEDEGGSGMGDGSSEFIQORFSETTEKYHAESIQNNNSKOELLIKEYELEKCLR	29
		:       :       :       :       :	
QY	301	MEENNRLRESKRLGGDDARVAELETLDRLRAEMLQLTTENNELHQOEAPRLSKFGD	359
		:       :       :       :       :	
Db	298	KEDENNRLRESKRLGGVDARVAELETLDRLRAEMLQLTTENNELHQOEAPRLSKFGD	356
		:       :       :       :       :	

RESULT 3  
Q8VDF7

ID	Q8VD7F	PRELIMINARY:	PRT:	246 AA.
AC	Q8VD7F			
DT	01-MAR-2002 (TReMBLrel. 20.	Created)		
DT	01-MAR-2002 (TReMBLrel. 20.	Last sequence update)		
DT	01-MAR-2002 (TReMBLrel. 20.	Last annotation update)		
DE	Similar to HMB4-inducible.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_Taxid:10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Strausberg R.;			
RL	Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: BC021111; AAH2111.1;			
SQ	SEQUENCE 246 AA; 28572 MW; 6CF8227F599B7EF CRC64;			
Query March	62.8%;	Score 1199;	DB 11:	Length 246;
Best Local Similarity	93.5%;	Pred. No.5.4e-78;		
Matches 229;	Conservative	5;	Mismatches 11;	Indels 0;
			Gaps	0;

OY	11	AELLAPRCIDSEASKIGAPAAAGCEEMWOOOLSKKKRRRRSSKKNPKRYUKLTWE	174
Dd	2	SSELVGRGHDSEVTQOEARAAGSERWGOOROLSKKKRRRRSSKKNPKRYUKLTWE	61
OY	175	KKKFEKOSLRASRIAEEMFAKOPVARYNTTQFLMDNDHOEDRDLKTGLYSKRAAKSD	234
Dd	62	KKFDEFKOILRASRYVAEMFAKQOPVARYNTTQFLMDNDHOEDRDLKTGLYPKRAAKSD	121
OY	235	DTSDDDPFMEBGEDEBGSGDMGGDSFPLJORDPFSFYERHPTESLOJNMSKOELKEYTEL	294
Dd	122	DTSDDEFVEABEBOGSGDMGGDSFPLJORDPFSFYERYAESLOJNMSKOELKEYTEL	181
OY	295	EKLSTBMEDENNRLRESKRLOGDGAARVELLELDLRLAENLOLTLENELHRQOERAPL	354
Dd	182	EKLSTBKEDENNRLRESKRLOGDVAVARELELDLRLAENLOLTLENELHRQOERAPL	241
OY	355	SKFGD 359 	
Dd	242	SKFGD 246 	

RESULT 4	
Q96MH2	
ID	Q96MH2
AC	Q96MH2
DT	01-DEC-2001 (TrEMBLrel. 19, Created)
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE	CCNA FLJ32384 fis, clone SKM51000104, weakly similar to Homo sapiens
DE	HCXIM1 protein (Similar to putative).
OS	Homo sapiens (human)
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX	NCBI_TaxID=9606;
ON	[1]

RC TTSUSE-SKELETAL MUSCLE;  
RA Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Tak  
RA Arita M., Mutsaers H. K., Yuuki H., Hara H., Sugiyama T.,  
RA Otsuki T., Saito H., Wakamatsu A., Ishii S., Yamamoto J.,  
RA Kawal-Hilo Y., Saito K., Nishikawa T., Kimura K., Yamash  
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.  
RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzu  
RA Sugano S., Nagahata K., Masuo Y., Nagai K., Isogai T.,  
RT "NEO human cDNA sequencing project.";  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases  
LN [2]  
RN  
RP SEQUENCE FROM N.A.  
RC TTSUSE-EYE;  
RA Strausberg R.;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases  
RR EMBL: AK056946; BAB71319.1; -.

Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases  
EMBL; AK056946; BAB71319.1; -





RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RT Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:  
"The genome sequence of *Drosophila melanogaster*";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BERKELEY;  
RA Stapleton M., Broksstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nuncio J., Pacled J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.B., Rubin G.M., Celisner S.;  
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AE003705; AAF55085.1; -  
DR EMBL: AY051786; AAK93210.1; -  
DR FlyBase: FBgn0038251; CG3508.  
SQ SEQUENCE 349 AA; 38058 MW; 8597FC7AE08D0747 CRC64;

Query Match 10.9%; Score 209; DB 5; Length 349;  
Best Local Similarity 25.9%; Pred. No. 3.1e-07;  
Matches 68; Conservative 43; Mismatches 84; Indels 68; Gaps 10;

QY 85 LREGEKONGD-----DSSAGDFPPPAEPTPEAELLAQPCNDSEASKLGAAPAGEE 139  
DB 1 MAEAVNNESSGQORPLDSSGGGG-----ASGGGVAVGGGS 36  
QY 140 EMGQOQROLGKKRRRPSKKRHRKPYKLTWE---EKKRPEKOSLRASIRAEFMA 195  
DB 37 GMRKRRHRRGKS-KMOPKTKNHY-PQKRLDSTGAGATLEGNQONSKTKLYRSL- 93  
QY 196 KGQVAPYNTQFLMDHDOEERDLKTYLKSRAAKSDDTSDDFMEEGEGEDGSDGM 255  
DB 94 ---LVYNTNRFLEMHSE-----LHKDSDNCF---GSQTE----- 126

QY 256 GGDGSEFLQDSEETERYHTESLQNMKSKOELKEYLEKLSRMEDENNRRLSKRL 315  
DB 127 --DQVFLSKERSDVERARLERLETMSKOELTQECOMJEDRYSKAONISKEP---GAKL 181

QY 316 GGDARVRELELELDRLRAENLQ 338  
DB 182 RAQDDKIRQLSRRENQFLRTHLLR 204

RESULT 7  
Q8T6B4 PRELIMINARY; PRT; 1142 AA.  
AC Q8T6B4;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Non-transporter ABC protein AbcF4.  
GN ABCF4.  
OS Dictyostelium discoideum (Slime mold).  
OC Eukaryota; Eukaryota; Dictyostelida; Dictyostelium.  
OX NCBI\_Taxid=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AX4;  
RA Anjard C., Loomis W.F.;  
RT "Evolution of the ABC transporters of *Dictyostelium*";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF479256; AAL87694.1; -  
SQ SEQUENCE 1142 AA; 130224 MW; 33B1815AB09942DC CRC64;

Query Match 7.9%; Score 150; DB 5; Length 1142;  
Best Local Similarity 19.8%; Pred. No. 0.02;  
Matches 79; Conservative 73; Mismatches 132; Indels 114; Gaps 18;

QY 30 PERPPAEERVPEDSRMOSRAFPQ---LGRPGPEGESLESQ-----PPLOTQA 78  
DB 132 PQKGGKQGO--QODSDQDEELIPQPVKGGKRAPOKGGKQODSDDEDELIPQVKKGG 189  
QY 79 CPSSSCLRGEKGQNGDSSAGGPPPA-----EVEPTPEAELLAQPC 122

DB 190 KPAPO--KKGGKOQSEDEDEDEVOQPVKGGKNDKKKGVKHVEEEREEIEQVY 247  
QY 123 HDEASKLGAAPAG-----EEENGQOQOLGKKRRRPSKKRHRKPYKLTWE 173  
DB 248 --KGGKAPKPKGGKSGKQSEDEEDDVOQPVKGGKDKKKGNH-----E 294  
QY 174 EKKRPEKOSLR-----ASIRAEFMAKQOPAPYNTQFLMDHDOEERDLKTYL- 224  
DB 295 EEEEEEERIEEDPVKGGSKKKQKGGKKNHVE-----EEEEEEIEQPVKGG 346  
QY 225 ---YSKRAAKSDTSDDFME-----EGGEEDGSDGSGDSEFLQDSEFYER 273  
DB 347 SNKKDQKGGKQOQOQSEDEDEEIEQPVKGGKDKKKK-----GSKHVEEEREE 400  
QY 274 YHTESLQNMKSKOELKEYLEKLSRMEDENNRRLSKRLGGDDARVRELE- 328  
DB 401 EEIE-----QPVKGGKKD-KSLSDSMSELSIKSK-GGKGNHVEEEREEQDEE 450  
QY 329 -----LDRLRAENLQTLTENELHROERAPLSK 356  
DB 451 EKKRKSNNKDKKGNHVE--EEEEEEEEEKPKSK 486

RESULT 8  
Q9NTH6 PRELIMINARY; PRT; 992 AA.  
AC Q9NTH6;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
DE Hypothetical 112.6 kDa protein (Fragment).  
GN DKEFZP43402413.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-TESTIS;  
RA Koehler K., Meyer A., Mewes H.W., Gassenhuber J., Wiemann S.;  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL137265; CAB70664.1; -  
KW Hypothetical protein.  
FT NON\_TER  
SQ SEQUENCE 992 AA; 112628 MW; 039DF5B1E97EF02F CRC64;

Query Match 7.8%; Score 148.5; DB 4; Length 992;  
Best Local Similarity 21.1%; Pred. No. 0.022;  
Matches 94; Conservative 66; Mismatches 149; Indels 137; Gaps 17;

QY 25 OEELNPERPPGAEEVP-----EEDSRMOSRAF-----POLGPRGPEG- 64  
DB 230 QDELQSKQSGLEERLSPPLPHERRAOSPPRSILATEEPPQ-----GPGQPEWKADEL 284  
QY 65 ---GSLESQPEPILOTA-CPSSSCLRGEKGQNGDSSAGGPPPAE---VEPP- 113  
DB 285 GEDSAASLSLQSLQREQAPSPAPACEKGEQNSQAEELGPGQGEAEDEPEKVAVSPTRP 344  
QY 114 -----EAELLAQPCNDSEASKLGAAPAGGEEEMQOQOLGKK----- 152  
DB 345 VSPERVSTPEVAPPEQLSEAA-LKAMEEVAQVLEDDQNHLLSKQEKMQOLREKLCQEE 403  
QY 153 -----HRRRPSKKRHRKPYKLTWEKKKFKDEQOSLRASIRAEFMAKQOPAPYNT 205  
DB 404 EBEITLRLHQQKQESLRLERLQKATEEEDARRREESQRLSLRAQVOSSTQA----- 457  
QY 206 TOFLMDHDOEERDLKTYLKSRAAKSD-----DTSDDFMEEGEGEGGS----- 252  
DB 458 -----DEQIARQDASLQKRLREELSSQOKAERASLEQKNQMLQLELLEASGSKSO 511  
QY 253 -----DQMGGDGSEFLQDSEFYER-----HTE---SLQNMK 284

```
Db 512 AALNAKAKALQOLREOLEGERKAVATLEKEHSALEERLCSSLEAKHREVSSLOKKIQ 571
OY 285 QELIKEYLELEKCLSRMEDENNRLRLLESKRRLGDDAVREL-----ELELDRL 332
Db 572 FAQOKEFAQLOKCLQGYE---HRYHOKSYHAGYEHHELSSLLREKROVEGEHERLDKM 628
OY 333 RAENLOLLT-----ENELHROQERAPL 354
Db 629 KEHOOVAKARQOYEAERKORAEI 654

RESULT 9
O9UPV0 PRELIMINARY: PRT: 1455 AA.
AC O9UPV0:
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE KIA1052 protein.
GN KIA1052.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE=99397452; PubMed=10470851;
RA Kikuno R., Nagase T., Ishikawa K., Hirose M., Miyajima N.,
Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 6:197-205(1999).
DR EMBL: AB028975; BAA83004.1; -
DR InterPro: IPR001202; WW_Rsp5_WWP.
DR Pfam: PF00397; WW; 1.
DR SMART: SM00456; WW; 1.
DR PROSITE: PS50020; WW_DOMAIN_2; 1.
SQ SEQUENCE 1455 AA; 163543 MW; 7F48093100C34819 CRC64;

Query Match 7.7%; Score 147; DB 4; Length 1455;
Best Local Similarity 21.6%; Pred. No. 0.044;
Matches 97; Conservative 65; Mismatches 146; Indels 142; Gaps 19;

OY 25 QEELNPEPPGAEEER-----VPEEDSRMOS--RAF-----POLGGRPEGEF 64
Db 456 QDELQSKOSKLEREYHRLSPPLPHEE-RAQSPPSRLATEEPPO-----GREGQPEKE 509
OY 65 -----GLESQPPPLQTA-CPESSCLREGKEKGONGDDSSAGDPFPPPAE---VE 110
Db 510 AEELGEDSAASLSLQSLRQAPSPAPACEKKGKHSQAELGQGEAEADPEKXAVS 569
OY 111 PTP-----EAELLAQPCHSKASKLAPAGGEFEEMGOOROLGKKK----- 152
Db 570 PTPVPSPEVRSSTEPAPPEQLSEAA-LKAMEEAAVAQVLEDDQRIHLESKMOOLREKL 628
OY 153 -----HRRRPSKKRHHMKPYUKLTWEKKKKPFDEKOSLSRASRIAEAFMAFGQVVA 201
Db 629 COEEBEELTRLHQKEQSLSLRELRLOKAITEEERKMEESQRLSMIRAAVOYSTQA-- 686
OY 202 PYNTTQFLMDHDDEPDLTGTLYSKRAAKSD-----DTSDDPFMEGEGEDGS- 252
Db 687 -----DEDDIRAQGEASLQKLRELEESQOKAEKASLEQKNRQMLDQLKEIEASE 736
OY 253 -----DGMGDGSEFLORDSETERY-----HTE---SLQ 280
Db 737 KSEQAAINAAKAKALQOLREOLEGERKAVATLEKEHSALEERLCSSLEAKHREVSSLO 796
OY 281 NMSKOELIKEYLELEKCLSRMEDENNRLRLLESKRRLGDDAVREL-----ELE 328
Db 797 KKIQDAQOKEFAQLOKCLQGYE---HRYHOKSYHAGYEHHELSSLLREKROVEGEHERR 853
```

```
OY 329 LDRLEAENLOLLT-----ENELHROQERAPL 354
Db 854 LDKMKHEHQVAKARQOYEAERKORAEI 883

RESULT 10
O95LS7 PRELIMINARY: PRT: 560 AA.
AC O95LS7:
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Hypothetical 63.1 kDa protein.
OS Macaca fascicularis (Crah eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxId=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-TESTIS;
RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
Terao K., Sugano S.;
RT "Isolation of novel full-length cDNA clones from macaque testis cDNA
RT libraries."
RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB071115; BAB64509.1; -
KW Hypothetical protein.
SQ SEQUENCE 560 AA; 63140 MW; C3EFD09DC4635E36 CRC64;

Query Match 7.5%; Score 142.5; DB 6; Length 560;
Best Local Similarity 22.4%; Pred. No. 0.03;
Matches 88; Conservative 55; Mismatches 121; Indels 129; Gaps 19;

OY 27 ELNPEPPGAEEERVPEDSRMOSRAFPQ-----LGRPGEGESLESQPPPLQ 75
Db 44 ESRPEEEEEEEREEESASOGTAADQAKVPKELTAFAAGEGEP-GEPGRAPKQPE- 101
OY 76 TQACPESSCLREGKGGONGDSAGGDP-----PRAVEYPTPEALLAOPCHNSEA- 127
Db 102 -----PE-----EPAEAGAEPPPOPKGAGPEELDAEARAEBELQAAEGEVR 144
OY 128 -----SKLG-----APAA-----GGEEMGOOROLGKKHRR-----RPSKKKHHWK 165
Db 145 SQASLPRLRIGEEAANAAPAEATERVGESEDEETRRDAGSESGRAGEGRAPKASQEGEK 204
OY 166 P-----YKLTWEKKKPFDEKOSLSRASRIAEAFMAFGQVAPYNTTQFLMDHDDEPD 219
Db 205 PLGRDFEDELWSE-----EVQKLOEQOLRSDDLQYR-----SLMEHNRSO--- 248
OY 220 LKTGLY-----SKRAAKSDPTSDDDPFMEGEGEDGS---SDGMGGGSGSEFLQORDF 267
Db 249 -RYNLYLDQHKLFELRKKKGLAEAEVDPKGAQAEAPKEQAYLRLHLM-----LEDLKKQ 303
OY 268 SETERYHTESLQMSKOELIKEYLELEKCLSRMEDENNRLRLLESKR-----G 316
Db 304 ADDLQWYN-----QELGQLKQOQEKLSRVKEKEMBRQALKQVNVQAMGSCRMG 354
OY 317 GDDARVREL-----ELELDRLAENLOL 339
Db 355 GROAALREVEQILALEDKKERKMSAVRLNVOL 387
```

```
RESULT 11
O9W475 PRELIMINARY: PRT: 1192 AA.
AC O9W475:
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE CG3108 protein.
GN CG3108.
OS Drosophila melanogaster (Fruit fly).
```

```

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abri J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borrova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferraz C., Ferlita S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glöckner A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Houston D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Relbert K., Remington K., Saunders R.D.C., Scheller F., Shen H.,
RA Shue B.C., Siden-Kimms I., Simpson M., Skupski M.P., Smith T.,
RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtka R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003435; AAF6083.1; -
DR HSSP: P00730; 2CTC.
DR FLYBase: FBgn0029807; CG3108.
DR InterPro: IPR000834; Zn_carpodpept.
DR Pfam: PF00246; Zn_carpodpept. 1.
DR PRINTS: PR00765; CRBOXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
SQ SEQUENCE 1192 AA; 132088 MW; 72408D5C5D3D718E CRC64;

Query Match 7.4%; Score 141.5; DB 5; Length 1192;
Best Local Similarity 22.3%; Pred. No. 0.086;
Matches 80; Conservative 57; Mismatches 138; Indels 83; Gaps 17;
OY 4 PFLSEYGHQPTSNCTGAAGVDELNPPRPGAEERVPEDSRWSRAFPQLGCRPGEG 63
DB 252 PLNDELPRDESPATTESAV--EELKESAMADQVPESEIPEQVQ---GEYSES 306
OY 64 EG-----SLESQPPPLQTOACPESSCLREGEKGQNDSSAGCDFPPAFAVEPTPEAE 116
DB 307 DGEQAEKPELEAQP---EVAQPEAEKQPEAE-----PQLEVEPEPEVE 348
OY 117 LLAQPCHDSEASKIGAPAGGEEEMGGQOQOLGKKHRRRPPYKRLTWEEK 176
DB 349 --SQPEVESQPEVAQPEVEPEVESQ-----PEASHSEPE---TQAEVE 390
OY 177 KFDKOSLRASRIAEKPAKQOPVAPYNTTQFLMDHDHQQEEDPLKTGLYSKRAAKSDT 236

```

```

DB 391 AQPEVESLPEAESQPE--AESQP-----ERPEVE---AEKISDNEVDY 429
OY 237 SDDPFMEGGEEDGSDGCGSEFLQDRDSEYTERHTESLQMSKQELIKYELEK 296
DB 430 TEASCLMEFLVE--GLEDGLTAMQNLVPEELAESDQKQETE-LESEDDQSPVTEAIE-EQ 485
OY 297 CLSRMEDENNRLRESKRLGCDGARVRELELDRLAENLQ-LTLENELRQOERAP 353
DB 486 AVPELEQEKER---EPQITLAD-----ETEDQASQPSNEPEVLAIEQHTAEINAP 534

RESULT 12
ID 076153 PRELIMINARY; PRT; 443 AA.
AC 076153;
DT 01-NOV-1998 (TREMBLrel. 08. Created)
DT 01-NOV-1998 (TREMBLrel. 08. Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19. Last annotation update)
DE Rsp60.
OS Periplaneta americana (American cockroach).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC Blattodea; Blattidae; Periplaneta.
OX NCBI_TaxID=6978;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20193551; PubMed=10727895;
RA Arai T., Kubo T., Natori S.;
RT "Identification, characterization and cDNA cloning of two novel
RT proteins secreted into the external space of the regenerating leg of
RT periplaneta americana."
RL Insect Biochem. Mol. Biol. 30:287-295(2000).
DR EMBL: AB012307; BAA32795.1; -
SQ SEQUENCE 443 AA; 47987 MW; 91AF646B161C99H CRC64;

Query Match 7.4%; Score 140.5; DB 5; Length 443;
Best Local Similarity 20.4%; Pred. No. 0.031;
Matches 80; Conservative 66; Mismatches 140; Indels 107; Gaps 17;
OY 4 PFLSEYGHQPTSNCTGAAGVDELNPPRPGAEERVPEDSRWSRAFPQL 55
DB 47 PVAAEVRKREPVAATAATYAAAVPKEDKPAEEVPPAPVVEEKKPAEDA----- 96
OY 56 GGRQPEGEGSLESQPPPLQTOACPE--SSCLREGEKGQND-----SSAGCDFPPA 107
DB 97 ---PSPAEAPAEQEVAP--DEAVPEVAQDETNNKGEVNDQSSDLKTATSSQDVAKPV 152
OY 108 E-----VEPTPAELLAOCHSEASKIGAPAGGEEEMGGQOQOLGK 150
DB 153 EEKVALNAPKVGEEPAEKVEEPAEKVEEVVQDQAAPTTAAEPKAEDEEPARDE--K 209
OY 151 KKHRRRP-SKKRRHMKPYKRLTWEEKKKKFDKOSLRASRIAEKPAKQOPVAPYNTQFL 209
DB 210 VEEAAPVSRKKRETAP-----KEEEKKPTTKKA-----KAE----- 242
OY 210 MDHDHDEPDRLKTGLYSKRAAKSDDTSDDPFMEGGEEDGSDGCGDSEFLQDRPSE 269
DB 243 -DVAVQEPKQDQVAVQEEVEVAVQEPKQDQVQDDAKKE-----EVKQEGDAKE 291
OY 270 TYEKVHTESLQMSKQELIKYELELKLCLSMEDENNRRLRESKRLGCD--DARVELE 326
DB 292 EEVAVQEEVADKADAAQVQKQDDV-----KVQDE--VKLEEVKVESDAKEEVKVEAK 341
OY 327 LELDRLAENLQLTLENELIRQ--OERRAPLSK 356
DB 342 VEDQAKVQEDVKV--QDEVKQDAAAEVNPVSK 372

```

```

RESULT 13
O8TXA4 ID O8TXA4 PRELIMINARY; PRT; 609 AA.
AC O8TXA4;
DT 01-JUN-2002 (TRFMBLrel. 21. Created)

```

DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Uncharacterized protein.  
 GN MK0771.  
 OS Methanopyrus kandleri.  
 OC Archaea: Euryarchaeota; Methanopyrii; Methanopyrales; Methanopyraceae;  
 OC Methanopyrus.  
 OX NCBI\_TaxID=2320;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AV19 / DSM 6324 / JCM 9639;  
 RX MEDLINE=21927647; PubMed=11930014;  
 RA Slesarev A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,  
 RA Shcherbdina O.V., Shakhova V.V., Belova G.I., Aravind L.,  
 RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,  
 RA Matlykh A.G., Koonin E.V., Kozayavkin S.A.;  
 RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19  
 RT and monophyly of archaeal methanogens";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).  
 DR EMBL: AE010369; AAM01985.1; -;  
 KW Complete proteome.  
 SO SEQUENCE 609 AA; 69552 MW; AB10C9780DC5AD78 CRC64;

Query Match 7.3%; Score 140; DB 17; Length 609;  
 Best Local Similarity 22.0%; Pred. No. 0.049;  
 Matches 80; Conservative 66; Mismatches 147; Indels 70; Gaps 17;

OY 31 ERPGAER-----VPEEDSRMOSRAFPOLGSRPEEGGS-----LESQPPPLQTQ 77  
 DB 5 KRKGKRRPDLTKLPKEGEGAPKLTLPKPEGRPG-ESKGALVAPKLTLPKPP----- 58  
 OY 78 ACPESSCLREGEKQCGDSSAGDPPEPAVE--PTPAELIAQCHDSKSLGAPAG 136  
 DB 59 -----KPSKPPSGSEEEKKEERPPALIKPPKEPRAPSNLEALILKAENKK 110  
 OY 137 GEE---EW-----GQOQOLG-KKRRRPPSKKKRHKMPYTKLTWEKKKDE---- 180  
 DB 111 LREELDEWRNKAASAMGERDLRSEIKRLEKELEKQELDKYIKISKOLKEKLEKARE 170  
 OY 181 ----KQSLRASRIRAFEMFAKGPVAPYNTQFLMDHDOEEDPKTGGLYSRAAKSDDT 236  
 DB 171 SEELKEKKEEYRERERKTA-----GKYNELSKSLDLSDONRLAENL--KKLEKYEI 223  
 OY 237 SDD-DFMEEGEEDGSDGMDGSEFLQROFSEF-YER---YHTESLQNMN---KOEL 287  
 DB 224 KEERDLKEEFKEVGKTK----DQAKLQSKLKEVKSRLDLANEVALRNEENELRKKI 279  
 OY 288 IKEYLELEKCLSRMEDENNRLLESKRLGDDARVRELELDRLRAENLQLLTENELHR 347  
 DB 280 DKLSELSNLOKKLDRKKLEKAROHQIKLREELIKRDEETRLKAQSKL--KDEIKR 337  
 OY 348 QOE 350  
 DB 338 YEF 340

RESULT 14  
 O94DE3 PRELIMINARY: PRT: 714 AA.

AC O94DE3.  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE P0683F02.9 protein.  
 GN P0683F02.9.  
 OS Oryza sativa (Rice)  
 OC Eukaryota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaeae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPONBARE;

RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(ca3) genomic DNA, chromosome 1, PAC  
 RT clone:P0683F02.";  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP003289; BAB63706.1; -;  
 SO SEQUENCE 714 AA; 77901 MW; 3A99C4DCA9155299 CRC64;

Query Match 7.3%; Score 140; DB 10; Length 714;  
 Best Local Similarity 21.8%; Pred. No. 0.06;  
 Matches 74; Conservative 52; Mismatches 132; Indels 82; Gaps 12;

OY 31 ERPGAERVPEEDSRMOS-RAFPOLGSRPEEGGSLESGPPLQTQACPESSCLREGE 89  
 DB 28 ERADOLAKRVALEEOVAVTAQRQARARATEVIGLIEHSQFQGNLSDVLDSGDRGE 87  
 OY 90 KG-----QNGDSSAGDPPPAPEVPTPEALIAQCHDSKSLGAPAGAEERWG 142  
 DB 88 EDDDPDRASDGDAGSGEGEOPPAO-----CEAEDALSGTAEPGGLSW- 133  
 OY 143 QOOROLGKKHRRRPPSKKKRHKMPYTKLTWEKKKFKDEKQSLRASRIRAFEMFAKGPAP 202  
 DB 134 -KGRSVSPRKARQLKQKIRRSY--FYLLSSDPSRYRMGQGRKK-RKEL-SNCKSTAP 188  
 OY 203 YNTQFLMDHDOEEDPKTGGLYSKRAAKSDDTSDDPMEEGEEDG-----SDGM 255  
 DB 189 -----EEOGRDVEELIAMSQKGOQDSDCTDDGADMDGCVAVGYVIRYKDG- 236  
 OY 256 GGDGSEFLQROFSEFYHTESLQNMNKKQELIKEYLELEKCLSRMEDENNRLLESKRL 315  
 DB 237 -----EMERYLER-----QAEILQYEAEEBAQORWEMKOPENNRSASAK-- 274  
 OY 316 GDDARVRELELDRLRAENLQLLTENELHROOERAPLS 355  
 DB 275 -----VHVEAKKACQIENGMSKSHSLA 300

RESULT 15  
 O15087 PRELIMINARY: PRT: 1520 AA.

AC O15087.  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE KIAA0383 protein (Fragment).  
 GN KIAA0383.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN.  
 RX MEDLINE=97349984; PubMed=9205841;  
 RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,  
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. VII.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro.";  
 RL DNA Res. 4:141-150(1997).  
 DR EMBL: AB002381; BAA20837.1; -;  
 DR InterPro: IPR002717; MOZ\_SAS.  
 DR InterPro: IPR001965; ZnF\_PHD.  
 DR Pfam: PF01853; MOZ\_SAS; 1.  
 DR SMART: PF00628; PHD; 1.  
 DR SMART: SM00249; PHD; 1.  
 FT NON\_TER 1  
 FT SEQUENCE 1520 AA; 171095 MW; BDC666120B0ADFEF2 CRC64;

Query Match 7.3%; Score 139; DB 4; Length 1520;  
 Best Local Similarity 18.6%; Pred. No. 0.17; Indels 150; Gaps 16;  
 Matches 84; Conservative 67; Mismatches 150; Indels 150; Gaps 16;

OY 7 SEYHOPOTSNGTG-----AAVQELNPPRPGAERVPEEDSRMOSRAFPOLG 56





JOURNAL  
COMMENT

Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
 Tissue Procurement: DCTD/DTF/Gazdar  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAMA13322 row: k column: 07  
 High quality sequence stop: 644.

```

FEATURES
source
Location/Qualifiers
1. .903
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6058110"
/clone_1lb="NH_MGC_68"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pCMV-SPOrt6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT Average insert size 1.8 Kb. Library constructed by Life Technologies."

```

BASE COUNT	21/ a	259 c	309 g	115 t	3 others
ORIGIN					

Alignment Scores:

Pred. No.	3,99e-92	Length:	903
Score:	1377.50	Matches:	251
Percent Similarity:	97.688	Conservative:	2
Best Local Similarity:	96.914	Mismatches:	4
Query Match:	69.504	Indels:	1
DB:	14	Gaps:	1

US-09-972-758A-2 (1-359) x BQ221641 (1-903)

QY	5	AlaPheProGluLeuLysGlyIleArgProGlyIleProGlyIleGlyIleGlyIleSerLeuLysGln	70
Db	1	GGTTTCCCCAGTTGGTGCGCTCCGGGGCCGGAGGGGAAGGAGCTCGAATCCCA	60
QY	71	ProProProLeuGlnThrClnAlaCysProGluSerCysLeuArgGluGlyIleLys	90
Db	61	CCACCTCCCTTCCAGCCCCAGGCGCTGCAGAAATCTAGCTGGCTTAGAGAGGGCCAGAG	12
QY	91	GlyGlnAsnGlyAspAspSerSerIleGlyIleAspPheProProAlaGluValGlu	11
Db	121	GGCCAGAAATGGGGAGAGACTCTCCCTGGCGGGCGACTTCCGCGCCGGCCAGAGTGGAA	18
QY	111	ProThrProGluAlaGluLeuLeuAlaGlnProCysSHAspSerGluAlaSerLysLeu	13
Db	181	CCGAGCGCCGAGGGCCGAGCTGCTCCGCCACGCTTCTGATCTGCAGGCGCAGTAACTTG	24
QY	131	GlyAlaProAlaAlaGlyIleGlyIleGluGluTrpGlyGlnGlnArgGluLeuGlyLys	15
Db	241	GGGGCTCCTCGCCGAGGGGGCCAAAGAGAGTGGGACAGCAGACAGACAGCTGGGGAG	30
QY	151	LysLysSHAspArgArgProSerLysLysLysArgHisTrpLysProTyrTyrLysLeu	17
Db	301	AAAAAACATAGAGAGCGCCCGTCCAGAAAGAGCGGCTTGGAAACCTTACTACAACTG	36
QY	171	ThrTrpGluGluLysLysPheAspGluLysGlnSerLeuArgAlaSerArgLysArg	19
Db	361	ACCTGGGAGAGAAATAAAAGTTCCAGCAAGAAACAGACCTTCGAGCTTCAAGATCCGA	42
QY	191	AlaGluMetPheAlaLysGlyIleProValAlaProTyrAsnThrThrGluPheLeuMet	21
Db	421	GGCGAGATGTTGCCAAGGGCCAGCGGGTGGCGCCCTTAACACCAACCACTTCTCTATG	48
QY	211	AspAspHisAspGlnGluProAspLeuLysThrGlyLeuTyrSerLysArgAlaAla	23
Db	481	GATGTATCAGACAGAGAGCCGATCTTCAAAACCGGCTGTACTCCAAACGGGGCCGC	54

OY	23	AlIySerAspAspPhrSerAspAspPheMetIuGIuGIuGIuGIuAspGly	250
Db	541	GCCAAATCCGACACACACGCATGACACTTCATGAAAGAGGGGGTGAAGAGAAAGG	6000
OY	251	GlySerAspGlyMetCylGlyAspGlySerGluPheGluGlnArgAspPheSerGluThr	2700
Db	601	GGCAGCGATGGCATGGAGCGGAGCGGACGCGATGTTCTGCAGCGGGAACCTCTCCGAAACG	6600
OY	271	TyrGluArgTyrHisThr-GluSerLeuGlnAsnMetSerLysGlnGluIuIuLysGln	2900
Db	661	TACGACCGGTACACACCGGAGAGCCCTCAGCAACATGTGAGCAAGCGAGGCTCATCAACAAG	7200
OY	290	uTyrluGluGlnLeuGlnLysCys---LeuSerArgMetSerGluAsnGlnAsnArg	307
Db	721	GTACCTGGGAGCTGAGAAATGGCCCTCTGCGCCGATGGAAGAGCAGAACCAACCGG	775

RESULT 2	AL552998	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS
AL552998	876 bp	mRNA	linear	EST 16-FEB-2001		
AL552998	LTI-NFL006-PL2	Homo sapiens	CD01072YL06.5			
	prime, mRNA sequence.					
AL552998.1	GI:12892417					
EST.						

FEATURES	Location/Qualifiers
source	1. .876

BASE COUNT	221 a	235 c	266 g	152 t	2 others
ORIGIN					

Alignment Scores:

Pred. No.:	1,74e-91	Length:	876
Score:	1319.00	Matches:	248
Percent Similarity:	99.208	Conservative:	0
Best Local Similarity:	99.208	Mismatches:	2
Query Match:	69.068	Indels:	1
DB:	9	Gaps:	0

US-09-972,758A-2 (1-359) x AL552998 (1-876)

QY	126	ATGGCGCAGACCCCTTTCTTGTGCACAAATATATACACACCGCTCCTCAACTAGCACTGTACAGGK	18
Db			
QY	21	AlaAlaAlaValaGlnGlnGlnIleuAsnProGluArgProProGlnValaGlnGlnIArgVal	40
Db			
QY	186	GGTGTGTGTGTACAGGAAGAGCGTGAACCTTAGCGCCCCCGAGCGCGGAGACGGGGT	245
Db			



QY 41 ProGUgluasPserArgrTrpGlnSerArGAlaPheProGUlnLeuGlyIaArProGU 60  
|||||  
246 CCCGAGAGAGACAGTAgGTGGCAATCGAGAGCGTTCCCGCAGTTGGGTGGCGCTCCGGG 305  
QY 61 ProGUglYglUGlySerLeuGlnSerGlnProProProLeuGlnThrGlnAlaCysPro 80  
|||||  
306 CCGGAGGGGGAACGAGCGCTGGAATCCCAACCCACTCCCTTGCAGACCCGCGCTGCCA 365  
QY 81 GlnSerSerCysLeuArgrGlnGlyIuGlyGlnAsnGlyAspAspSerSerAlaGly 100  
|||||  
366 GAATCTAGCTGCTGTGAGAGGGGCGAAGGGCCAGAAATGGAGCAGCTGCTCCGCTGGC 425  
QY 101 GlysPheProProProAlaGlnValGluProThrProGUlnaGlnLeuAlaGln 120  
|||||  
426 GGGGACTTCCCGCGCGCGGAGAGTGGAAAGCCAGCCCGAGGCGGAGCTGCGCCAG 485  
QY 121 ProCysHisAspSerGlnAlaSerLysLeuGlyAlaProAlaAlaGlyGlnGlnGlu 140  
|||||  
486 CCTTGTATGACTCCGAGGGCCAGTAAGTTGGGGGCTCTCCCGCAGC-GCGGAAGAGAG 544  
QY 141 TrpGlyGlnGlnGlnArgrGlnLeuGlyLysLysLysHisArgrArgrProSerLysLys 160  
|||||  
545 TGGGAGACGACGAGAGACAGCTGGGGAAGAAAACATAGAGACGCCCGCTCCAAAG 604  
QY 161 LysArGHisTrpLysProTrpTrpLysLeuThrTrpGlnGlnLysLysLysPheAspGlu 180  
|||||  
605 AAGCGGCAATTGGAAACCGTACTCAACGCTACCTCGGAGAGAGAAAGAAAGTTCCAGAG 664  
QY 181 LysGlnSerLeuArGlnAlaSerArGlnAlaGlnAlaGlnPheAlaLysGlyGlnProVal 200  
|||||  
665 AAACAGAGCGCTTCAGCTTCAAGATCCGAGCGAGATGTTCCCAAGGCGCCAGCGCTG 724  
QY 201 AlaProTrpArGrnThrGlnPheLeuMetAspAspHisAspGlnGlnGluProAspLeu 220  
|||||  
725 GCCCGCTATACACCGACGCTTCATGATGATACAGACAGGAGGAGCGCGAGATCTC 784  
QY 221 LysThrGlyLeuTrpSerLysArGlnAlaAlaLysSerAspAspThrSerAspAsp 240  
|||||  
785 AAACCGCGCTGTACTCTCAAGCGCGCGCGCCCAATCCGACGACACACGATGAGACAG 844  
QY 241 PheMetGlnGlnGlnGlyGlnGlnGly 250  
|||||  
845 TTTCATGGAARAGCGGCTGAGAGATGCG 874  
Db  
RESULT 3  
B1871190 771 bp mRNA linear EST 11-OCT-2001  
LOCUS B1871190 603395081F1 NIH\_MGC\_90 Homo sapiens cDNA clone IMAGE:5404724 5',  
DEFINITION mRNA sequence.  
B1871190  
ACCESSION B1871190.1 GI:16044865  
VERSION B1871190.1  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 771)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LHAM12032 row: b column: 21  
High quality sequence stop: 733.  
Location/Qualifiers

source 1. 771  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5404724"  
/clone\_lib="NIH\_MGC\_90"  
/tissue\_type="adenocarcinoma, cell line"  
/lab\_host="Dh10b (phage-resistant)"  
/note="Organ: liver; Vector: pCMV-Sport6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 1.7 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH-MGC library."  
BASE COUNT 187 a 213 c 273 g 98 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1.61e-87 Length: 771  
Score: 1266.50 Matches: 248  
Percent Similarity: 96.51% Conservative: 1  
Best Local Similarity: 96.12% Mismatches: 5  
Query Match: 66.31% Indels: 5  
DB: 13 Gaps: 1  
US-09-972-758a-2 (1-359) x B1871190 (1-771)  
QY 21 AlaAlaAlaValGlnGlnGlnGlnLeuAsnProGUlnArGrProGUlnaGlnArGVal 40  
|||||  
3 GCTGCTGCTCTCCAGAGAGAGCTGAACCTCGAGCGCCCGCGGCGGAGAGCGGGT 62  
QY 41 ProGUgluasPserArgrTrpGlnSerArGAlaPheProGUlnLeuGlyIaArProGU 60  
|||||  
63 CCGGAGAGAGACAGTAgGTGGCAATCGAGAGCGTTCCCGCAGTTGGGTGGCGCTCCGGG 122  
QY 61 ProGUglYglUGlySerLeuGlnSerGlnProProProLeuGlnThrGlnAlaCysPro 80  
|||||  
123 CCGGAGGGGGAAGGAGCGCTGGAATCCCAACACTCTCTTGAGACCCAGCGCTGTGCA 182  
QY 81 GlnSerSerCysLeuArgrGlnGlyIuGlyGlnAsnGlyAspAspSerSerAlaGly 100  
|||||  
183 GAATCTAGCTGCTGTGAGAGGGGCGAGAAAGGCCAGATGGAGCAGCTGCTGCGCTGGC 242  
QY 101 GlysPheProProProAlaGlnValGluProThrProGUlnaGlnLeuAlaGln 120  
|||||  
243 GCGGACTTCCCGCGCGCGGAGAGTGGAAACCGACCCCGAGCGCTGCTGCGCCAG 302  
QY 121 ProCysHisAspSerGlnAlaSerLysLeuGlyAlaProAlaAlaGlyGlnGlnGlu 140  
|||||  
303 CCTTGTATGACATCCGAGCGCAGTAAGTTGGGGCTCTCCCGCAGGCGCGGAAGAGAG 362  
QY 141 TrpGlyGlnGlnGlnArgrGlnLeuGlyLysLysLysHisArgrArgrProSerLysLys 160  
|||||  
363 TGGGAGACGACGAGAGACAGCTGGGAGAGAAAACATAGGAGAGCGCCGCTCCAAAGAG 422  
QY 161 LysArGHisTrpLysProTrpTrpLysLeuThrTrpGlnGlnLysLysLysPheAspGlu 180  
|||||  
423 AAGCGGCAATTGGAAACCGTACTCAAGCTACCTCGGGAACAGAAAGATTTCAGAGAG 482  
QY 181 LysGlnSerLeuArGlnAlaSerArGlnAlaGlnAlaGlnPheAlaLysGlyGlnProVal 200  
|||||  
483 AAACAGAGCGCTTCGACCTTCAGAGATCCGAGCGAGATGTTCCCAAGGCGCGCGCTG 542  
QY 201 AlaProTrpArGrnThrGlnPheLeuMetAspAspHisAspGlnGlnGluProAspLeu 220  
|||||  
543 GCGCGCTATACACCGACGCTTCATGATGATGATCAGCAGCAGAGAGCGCGGATCTC 602  
QY 221 LysThrGlyLeuTrpSerLysArGlnAlaAlaLysSerAspAspThrSerAspAsp 240  
|||||  
603 AAACCGCGCTGTACTCAAGCGGGCGCG-GCCAAATCCGACGACACCAAGCATGACGAG 661  
QY 241 PheMetGlnGlnGlnGlyGlnGlnGlyLysArGlnGlyLysArGlnGlyLysGly---GlysPglY 259  
|||||  
662 TTTCATGGAAGAGCGGCTGAGAGATGGGCGGCGAGCATGGAATGGGAAGCGCGACCGCC 721

RESULT 4	BP984049	982 bp	mRNA	linear	EST 23-JAN-2001
LOCUS	BP984049				
DEFINITION	260	serlu:linarglnarasp-phersegu-THIRYARGUARGTyr	274		
	722	AGCAGAGGTTCTCCAGCGGAACTTCTTGAGAAACGTACGACCGGTAC	769		
ACCESSION	BP984049				
VERSION	BP984049.1				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 982)				
TITLE	NIH-MGC http://mgc.ncl.nih.gov/.				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-re@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <a href="http://image.lnl.gov">http://image.lnl.gov</a> plate: LLM10100 row: a column: 23 High quality sequence stop: 691. Location/Qualifiers 1. .982 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4398238" /clone_lib="NIH-MGC-88" /tissue_type="duodenal adenocarcinoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: small intestine; Vector: pCMV-SPORT6; site:1: NotI; Site:2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.767 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH-MGC Library."				
BASE COUNT	261 a	235 c	320 g	166 t	
ORIGIN					
Alignment Scores:					
Pred. No.:	1.05e-84	Length:	982		
Score:	1231.00	Matches:	238		
Percent Similarity:	99.17%	Conservative:	0		
Best Local Similarity:	99.17%	Mismatches:	2		
Query Match:	64.45%	Indels:	1		
DB:	12	Gaps:	0		
US-09-972-758A-2 (1-359) x	BP984049 (1-982)				
QY	120	GlnProCysHisAspSerGlnAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGlnGlu	139		
Db	8	CAGCCTTGTCATGACTCCGAGCCAGTAAAGTTGGGGCTCTCCGCGCAGGGGCGAAGAG	67		
QY	140	GluTrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLys	159		
Db	68	GAGTGGGGACGCGAGCAGACGTGGGGAAGAAAAACAATAGGAGAGCGCCGTCACAG	127		
QY	160	LysLysArgHisTrpLysProTyrTyrLysLeuThrTrpGlnGlnLysLysLysPheAsp	179		
Db	128	AAGAAGCGGCCTTGGAAACCTACTACAGCTGCACTCGGAAGAGAAACAAAAGTTCCAC	187		
QY	180	GluLysGlnSerLeuArgAlaSerArgTlleArgAlaGluMetPheAlaLysGlyGlnPro	199		
Db	188	GAGAAACAGACCCCTTCGAGCTTCAAGGATTCGAGCGCGAGATGTTCCGCAAGGCGACGCG	247		

	QY	200	ValAlarProtyrAsnThrThGlnPheLeumetKspSPHtSASPgInGlUpRoasp	219
	Dd	248	GTCGCCCGCCCTTAAACACCAGCACTTCTCTGTATGATCAGCACGAGAGACCGGAT	307
	QY	220	LeuLysThrGliLeuTrSerLysArgAlaIalAlySerAspSpPrThSerAspSp	239
	Dd	308	CTCAAACAACGGCCCTGTACTCCAACCGGGCCGCCCAATCTCCAGCAGCACACCGCATGCAC	367
	QY	240	AspPheMetGluGluGlyGlyGlnGluAspGlyLysSerAspGlyMetGlyGlyAspGly	259
	Dd	368	GACTTCATGGAAGAAGGGGGTGAGAGAGATGGCGCACCGATTGGGAATGGAGGGGACGGC	427
	QY	260	SerGluPheLeuGluGlnAspSPHeSerGIUThrTryGluArqTyRHtStHGluSerLeu	279
	Dd	428	AGCAGATTCTTCGACGGCGGACTTCTCGAGACGTACAGCGCGGTACCAACAGCAGACGCTG	487
	QY	280	GlnAsmetsrLyrsIngIngluLeuIllelsglsutryrLeugluGluLysCylser	299
	Dd	488	CAGAACATGTGAGCAAGCAGACCTCATCAGAAGTAACTTGGAAACGTGAGAACTGCTCTCG	547
	QY	300	ArgmetGluASpGLySuAnsAnsrJeuArgeuleuInsertyrAsrgleuGlyLYAspAsp	319
	Dd	548	CGCATGTGAGCAGCAGAACACCGGCTCGCGTGTGAGACAAACGGCTCGGTGGCGACGAC	607
	QY	320	AlaArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnleu	339
	Dd	608	GCGCGTGTGGGGAGCTGTGAGACTGTGACCTGTGACCGCTGCCCGCGGAACCTCCACTCG	667
	QY	340	LeuthrGluAsngluLeuHisARGIingIngluArgAlaProLeuSetrysPhyeGlyVAsp	359
	Dd	668	CTGACCCAGAACGAACTGTCCACCGGACGAGAGAGCGCG - CTTCCCAGATTGGAGAC	726
	RESULT 5	BG819109	797 bp mRNA linear EST 22-MAY-2001	
	LOCUS	BG819109	60278120.F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4932025	
	DEFINITION	BG819109	5' , mRNA sequence.	
	ACCESSION	BG819109	BG819109 GI:14166696	
	VERSION	BG819109.1	EST.	
	KEYWORDS	Homo sapiens	human.	
	SOURCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
	ORGANISM	NIH-MGC http://mgc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC)	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapsb@mail.nih.gov Tissue Procurement: David N. Louis, M.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov plate: LLAM10857 row: C column: 02 High quality sequence stop: 797.	
	COMMENT	Location/Qualifiers		
	FEATURES	source	/organism="Homo sapiens"	
			/db_xref="taxon:9606"	
			/clone="IMAGE:4932025"	
			/clone_lib="NCI-CGAP_Brn67"	
			/tissue_type="neoplastic oligodendroglioma with lp/19q loss"	
			/lab_host="DH10B (TI phage-resistant)"	
			/note="Organ: brain; Vector: PCMV-SPOrt6; Site.1: NotI; Site.2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.3 kb. Constructed by Life Technologies. Note: a NCI-CGAP Library."	
	BASE COUNT	207 a 223 c 248 g 119 t		

## ORIGIN

## Alignment Scores:

Pred. No.: 1,27e-79 length: 797  
 Score: 1163.50 Matches: 232  
 Percent Similarity: 95.12% Conservative: 2  
 Best Local Similarity: 94.31% Mismatches: 8  
 Query Match: 60.92% Indels: 6  
 DB: 12 Gaps: 1

US-09-972-758a-2 (1-359) x BG819109 (1-797)

QY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20  
 |||||  
 Db ATGCCGAGACCTTCTTGTGCAGATATCAACACACAGCTCAAACTAGCAACTGTACAGGT 128  
 QY 21 AlaAlaAlaValGlnGlnGluLeuAsnProGlnArgProGlnAlaGlnGluArgVal 40  
 |||||  
 Db 129 GCTGCTGCTGTCCAGAGAGAGCTGAACCTCGAGCGCCCGCCAGCGCGAGAGCGGGTG 188  
 QY 41 ProGlnGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60  
 |||||  
 Db 189 CCCGAGAGAGACAGCTAGTGTCATCGAGAGCTTCCCGAGTGGGTGGCGCTCCGGGG 248  
 QY 61 ProGlnGlyGlnGlySerLeuGlnSerGlnProProProLeuGlnThrGlnAlaCysPro 80  
 |||||  
 Db 249 CCGGAGGGGGAAGGAGCGCTGGAATCCCAACACCTCCCTGCAGACCCAGCGCTGTCCA 308  
 QY 81 GluSerSerCysLeuArgGlnGlyGlyGlyGlnGlnGlnGlnGlnGlnGlnGlnGln 100  
 |||||  
 Db 309 GAATCTAGCTGCTCGAGAGGGCGGAGAGGGCCAGATGGGAGCAGCTCTCCGCTGCG 368  
 QY 101 GlyAspPheProProProAlaGlnValGlnProThrProGlnAlaGlnLeuLeuAlaGln 120  
 |||||  
 Db 369 GCGGACTTCCCGCGCGCGCGGAGAGTGGAAACCCGAGCGCCGAGCGTGTGGCCAG 428  
 QY 121 ProCysHisAspSerGlnAlaSerLysLeuGlyAlaProAlaAlaGlnGlyGlnGlnGln 140  
 |||||  
 Db 429 CCTTGTATGACTCCGAGGCGCAGTAGTTGGGGGCTCTGCCCGCAGGGGGCGAAGAGAG 488  
 QY 141 TrpGlnGlnGlnGlnArgGlnLeuGlnLysLysLysHisArgArgProSerLysLys 160  
 |||||  
 Db 489 TGGGGACACGACGAGAGACAGCTGGGGAAGAAAAACATAGGAGCGCCGCTCCAGAG 548  
 QY 161 LysArgHisTrpLysProTyrTrpLysLeuThrTrpGlnGlnLysLysLysLys 180  
 |||||  
 Db 549 AAGCGGCA-TGGAACCGTACTACAGCTGAGCTGGGAAGAGAAACAAAGCTTCGACG 607  
 QY 180 LysGlnSerLeuArgAlaSerArg11eArgAlaGlnLeuPheAlaLysGlnPro 200  
 |||||  
 Db 608 AGAAACAGAGCGCTTCAGGCTTCAAGATCCGAGCCGAGATGTTCCGCAAGGGCCACCGG 667  
 QY 200 AlaAlaProTyrAsnThrThrGlnPheLeuMet-AspAspHisAspGlnGlnProAsp 219  
 |||||  
 Db 668 TCGCGGCAATATACACACGAGCTGCTCATGGAGATCAGACACGAGGGAGCGCGAT 727  
 QY 220 LeuLysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAsp 239  
 |||||  
 Db 728 CTCGAACCGGGGCTG-TACTTCGAAGCGGGCGCGCA---AATCCGAGACACAGGATGAC 783  
 QY 240 AspPheMetGln 243  
 |||||  
 Db 784 GACTTCATCGAA 795

## RESULT 6

AV726891 750 bp mRNA linear EST 17-OCT-2000  
 LOCUS AV726891 HMC Homo sapiens cDNA clone HTCA0E07 5', mRNA sequence.  
 DEFINITION AV726891  
 ACCESSION AV726891  
 VERSION AV726891.1 GI:10836312  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

## REFERENCE

1 (bases 1 to 750)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## AUTHORS

Gu, Y., Peng, Y., Song, H., Huang, Q., Yang, Y., Gao, G., Xiao, H., Xu, X.,  
 Li, N., Qian, B., Liu, F., Qu, D., Gao, X., Cheng, Z., Xu, Z., Deng, L., Xu,  
 S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R.,  
 Chen, J., Chen, Z., and Han, Z.

## TITLE

Homo sapiens cDNA HTC clones

## JOURNAL

Unpublished (2000)

## COMMENT

Contact: Zeguang Han

Chinese National Human Genome Center at Shanghai  
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
 201203, P. R. China

Tel: 86-21-50801919 (ex.45)  
 Fax: 86-21-50801922

Email: hanzg@cnhc.sh.cn  
 This clone is available at CHGC in Shanghai.

## FEATURES

## source

1. 750  
 location/Qualifiers

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="HTCA0E07"

/clone\_lib="HTC"

/tissue\_type="Hypothalamus"

/dev\_stage="Adult"

/lab\_host="SOLR"

/note="Vector: pBluescript SK(-); site\_1: EcoRI; site\_2:  
 XhoI"

## BASE COUNT

184 a 208 c 251 g 96 t 11 others

## ORIGIN

## Alignment Scores:

Pred. No.: 5.24e-77 length: 750  
 Score: 1129.00 Matches: 219  
 Percent Similarity: 85.94% Conservative: 1  
 Best Local Similarity: 85.55% Mismatches: 18  
 Query Match: 59.11% Indels: 18  
 DB: 10 Gaps: 1

US-09-972-758a-2 (1-359) x AV726891 (1-750)

QY 101 GlyAspPheProProProAlaGlnValGlnProThrProGlnAlaGlnLeuLeuAlaGln 120  
 |||||  
 Db 2 GCGGACTTCCCGCGCGCGGAGAGTGGAAACCCGAGCGCCGAGCGTGTGGCCAG 61  
 QY 121 ProCysHisAspSerGlnAlaSerLysLeuGlyAlaProAlaAlaGlnGlyGlnGlnGln 140  
 |||||  
 Db 62 CCTTGTATGACTCCGAGGCGCAGTAGTTGGGGGCTCTGCCCGCAGGGGGCGAAGAGAG 121  
 QY 141 TrpGlnGlnGlnGlnArgGlnLeuGlnLysLysLysHisArgArgProSerLysLys 160  
 |||||  
 Db 122 TGGGGACGACGAGAGACAGCTGGGGAAGAAAAACATAGGAGAGCGCCGCTCCAGAG 181  
 QY 161 LysArgHisTrpLysProTyrTrpLysLeuThrTrpGlnGlnLysLysLysLys 180  
 |||||  
 Db 182 AAGCGGCAATGGAACCGTACTACAGCTGAGCTGGGAAGAGAGAG----- 226  
 QY 181 LysGlnSerLeuArgAlaSerArg11eArgAlaGlnLeuPheAlaLysGlnPro 200  
 |||||  
 Db 227 -----ATCCCAAGGGCCAGCGCTC 247  
 QY 201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGlnProAsp 220  
 |||||  
 Db 248 GCGCCTATATACCCAGCGAGTTCCTCATGATGATCAGACACGAGAGAGCGGAGCTTC 307  
 QY 221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAsp 240  
 |||||  
 Db 308 ANAACCGGCGTGTACTCAAGCGGGCGCGCCCAAAATCCGAGACACCGAGTGCAGC 367  
 QY 241 PheMetGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260  
 |||||  
 Db 368 TTTCATGGAAGAGAGGGGTGAGGAGATGNGGACGCGATGGAGTGGAGGGAGCGGACG 427

QY	261	GIuprhleuGlnAaGAspRheSergIuThrTyrgIuARqTYrHsIthGluSerIeuGln	280
Db	428	GAGTTTCGCGAGCGGAGACTTCTCGGAGACGTCAGACGCGGTACCAACAGGAGACTTCGAG	487
QY	281	AsmSetrLySgInGluIleuIlelSgIuTyrIeuGluIeuGluLyCysIeuSerArg	300
Db	488	AACATGACACAAGCAGAGACTCATCAAGAGATACCTGGAACTGGAAGATGCTTTCGCGC	547
QY	301	MetGluAspGluAsnAsnArgIeuArgIeuIuSerLySArgIeuGlyIAspAspAla	320
Db	548	ATGAGAGCGAGAACACCGGCTGCTGAGACCAAGCGGCTGGGTGGAGACGACCG	607
QY	321	ArgValArgIuIeuGluIleuGluIleuAspArgIeuArgIaGluAsnIleuGlnIleu	340
Db	608	CCTGGCGGAGCTTGAGCTGACCTTGACCTTGACCCGCTTGNNCCCGAGACCTTCAGTCTG	667
QY	341	ThrgIuAsnGluIeuHsIArgGlnGlnIuArgIuArgIuAroIeuSerLyS	356
Db	668	ACCGAANAAGAACCTGACCCGACGACANAGCGAGGCGGCTTCCAAAG	715
RESULT 7			
LOCUS	BI091005	852 bp	mrna linear EST 20-JUN-2001
DEFINITION	602853667/1 NIH_MGC_10 Homo sapiens CDNA clone IMAGE:4995065 5',		
ACCESSION	BI091005		
VERSION	BI091005.1	GI:14509335	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgsaps-remail.nih.gov		
	Tissue Procurement: ATCC		
	cDNA Library Preparation: Life Technologies, Inc.		
	cDNA Library Arrayed by: Incyte Genomics, Inc.		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNL at:		
	http://image.llnl.gov		
	Plate: LRAM1018	row: e	column: 18
	High quality sequence stop: 700.		
FEATURES			
source	location/Qualifiers		
	1..852		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4995065"		
	/clone_lib="NIH_MGC_10"		
	/cell_line="MGC36"		
	/lab_host="DH10B"		
	/note="Organ: cervix; Vector: pCMV-SPORT6; Site.1: NotI;		
	Site.2: SalI; Cloned unidirectionally. Primer: Oligo dT.		
	Average insert size 1.5 kb. Library prepared by Life		
	Technologies.		
BASE COUNT	238 a	196 c	132 t
ORIGIN			
Alignment Scores:			
Pred. No.:	2,88e-76	Length:	852
Score:	1120.00	Matches:	226
Percent Similarity:	98.70%	Conservative:	1
Best Local Similarity:	98.26%	Mismatch:	3
Query Match:	58.64%	Indels:	3
DB:	13	Gaps:	0
US-09-972-758A-2 (1-359) x BI091005 (1-852)			
QY	130	LeuGlyAlaProAlaIaGlyGlyGluGlnIuTrpGlyGlnGlnIuArgGlnIeuGly	149

Db	5	GTGGGGCTCTGCCCGACGGGGCGCAAGAGCTGGACAGCAGACAGACAGCTGGGC	64
Qy	150	LysLysLysHisAArgArgProSerLysLysLysAArgHisTPLeysProTyTyLys	169
Db	65	AAGAAAAAACATAGAGAGACGCCCGCTCCAAAGAAAGACGGCC - TGGAAACCTTACTACAG	123
Qy	170	LeuThrTPGInuGInuLysLysLysPheAspGluLysGlnSerLeuArgAlaSerArgLe	189
Db	124	CTGACCTTGGAGAGACAAAGAAAGTTCCAGCAGAAACAGACCTTCAGACTTCACAGATC	183
Qy	190	ArgAlaGluMetPheAlaLysGlyGlnProValAlaProTyArgSerThrGlnPheLeu	209
Db	184	CGACCCCGAGATGTTGCGCCAAAGGGCCAGCGGTGCGCCCTTAACACACACAGTCTCTC	243
Qy	210	MetAspAspHisAspGlnGluGlnProAspLeuSerThrGlyLeuTySerLysArgAla	229
Db	244	ATGATGATCATCGACACAGAGAGAGCGGATCTCAAAACCGGCCTGTACTCCAAAGGGGAC	303
Qy	230	AlaAlaLysSerAspAspThrSerAspAspPheMetGluGluGlyGluGlnLysP	249
Db	304	GCCTGCCTAATCCGACGACACACAGCATGACACTTCATGGAAGAAAGGGGTGAGAGGAT	363
Qy	250	GlyLysSerAspGlyMetGlyLysArgLysGlySerGluPheLeuGlnArgAspPheSerGlu	269
Db	364	GGGGGCAACCGATGGGATGCGGAGCGACCGCACCGCGTCTG - CAGCGGAGCTTCTCGGAG	422
Qy	270	ThrTyArgTyrGlyThrHisThrGluSerLeuGlnAsnMetSerLysGlnGluLeuLys	289
Db	423	ACGTACGACGGCGGTACACACCGGAGACCTTGCAGAAATGAGACCAAGAGAGACTCATCAAG	482
Qy	290	GluTyTyLeuGlnLeuGlnLysCysLeuSerArgMetGluAspGluAsnAsnArgLeuArg	309
Db	483	GAGTACCTCGGAACGTGAGAACATGCTCTCCGCGCATGCGAGACGAGAAACACCGCTCGGG	542
Qy	310	LeuGluSerLysArgLeuGlnGlyLysAspAspAlaArgValArgGluLeuGlnLeuGlnLeu	329
Db	543	CTGAGAGACAAAGCGCTGGGTGGGACACCGCGCTGTGGGGAGACTGGAACTGGAACTG	602
Qy	330	AspArgLeuArgAlaGluAsnLeuGlnLeuLeuThrGlnAsnGluLeuHisArgGlnGln	349
Db	603	GACCGGCTGCGCGCGGAGAACCTCCAGCTGTGACCGGAGAACGAATGTCACGCG - CAGCAG	661
Qy	350	GluArgArgAlaProLeuSerLysPheGlyAsp	359
Db	662	GAGCGAGCGCGCTTTCCAAAGTTTGGAGAC	691
RESULT 8			
BG291150			
LOCUS	871 bp	mRNA	linear
DEFINITION	60238718221 NIH_MGC_93 Homo sapiens CDNA clone IMAGE:4515812 5',		
ACCESSION	BG291150		
VERSION	BG291150.1	GI:13048811	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 871)		
JOURNAL	NIH-MGC <a href="http://mhc.nci.nih.gov/">http://mhc.nci.nih.gov/</a> .		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: <a href="mailto:cgabs-remail.nih.gov">cgabs-remail.nih.gov</a>		
	Tissue Procurement: ATCC		
	CDNA Library Preparation: Life Technologies, Inc.		
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLND)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLND at:		
	<a href="http://image.llnl.gov">http://image.llnl.gov</a>		
	plate:LLAM10406	row: d	column: 21



```

Db      123 CCACTCCCTTGCAACCCAGGCGCTGTCAGAAATCTAGCTGAGAGAGAGGCGGAGAG 182
QY      91 GYGLInsnGlyAspSerSerAlaGlyIAspPheProProAlaGluValGlu 110
Db      183 GGGCAGATGGGAGCGACTGCTCGCTGGGGGAGACTTCCCGCGCGGAGAGAGTGGAA 242
QY      111 ProThrProGluAlaGluLeuAlaGluProCysHisAspSerGluAlaSerLysLeu 130
Db      243 CCGAGCGCCGAGGCGCAGCTGCTGCGCCAGCTTGCATGACTCCGAGCCAGTAAGTTG 302
QY      131 GYAlaProGluAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 150
Db      303 GGGGCTCTGCTCCAGGGGCGAGAGAGAGTGGGAGAGAGAGAGAGAGAGAGAGAG 362
QY      151 LysLysHisArgArgArgProSerLysLysLysArgHisTrpLysProTyrTyrLysLeu 170
Db      363 AAAAAACATGAGAGAGCGCCGCTCCAGAAAGACGGCATTTGGAAACCGTACTCAAGCTG 422
QY      171 ThrTrpGluGluLysLysLysPheAspGluLysGlnSerLeuArgAlaSerArgIleArg 190
Db      423 ACCTGGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 482
QY      191 AlaGluMetPheAlaLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 210
Db      483 GCGAGAGATGTTCCCAAGGCGCAGCGGCTGCGCCCTATACACACAGAGATTCTCAT 542
QY      210 TAspAspHisAspGluGluGluProAspLeuLysTrpGlyLysLysArg-Ala 230
Db      543 GGTATGATACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 602
QY      230 La-AlaLysSerAspAspThrSerAspAsp-----AspPheMetGluGlyGly 246
Db      603 CCGCGCAATCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 662
QY      247 GlnGluAspGlyGly---SerAspGlyMetGlyGlyAspGlySerGlu----- 261
Db      663 AAGAGAGATGGGGGCGGCGGAGTGGGAATGGGGAAGGGGAAGCGCGCGCAATTTT 722
QY      262 PheLeuGln 264
Db      723 TTCTTCAG 731

RESULT 10
AL578242/c 1001 bp mRNA linear EST 16-FEB-2001
LOCUS      AL578242 LTI_NFL006_P12 Homo sapiens cDNA clone CS0DK002YF07 3
DEFINITION prime, mRNA sequence.
ACCESSION AL578242
VERSION    AL578242.1 GI:12942132
KEYWORDS   EST.
SOURCE      human.
ORGANISM   Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1001)
AUTHORS     Li, W.B., Gruber, C., Jessee, J. and Polyes, D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 Evry cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
            Location/Qualifiers
                1..1001
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="CS0DK002YF07"
                /clone_lib="LTI_NFL006_P12"
                /issue_type="placenta"
                /note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA
                enriched, double-stranded cDNA was digested with Not I and
                cloned into the Not I and Eco RV sites of the pCMVSPORT 6

```

```

vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com
BASE COUNT      218 a      289 c      218 g      271 t      5 others
ORIGIN
Alignment Scores:
Pred. No.:      3 8e-72      Length:      1001
Score:          1067.00      Matches:      209
Percent Similarity: 97.66%      Conservative: 0
Best Local Similarity: 97.66%      Mismatches: 5
Query Match:     55.86%      Indels:      1
DB:              9          Caps:      0

US-09-972-758A-2 (1-359) x AL578242 (1-1001)
QY      146 ArgGlnLeuGlnLysLysLysHisArgArgArgProSerLysLysArgHisTrpLys 165
Db      999 AGACAGCTGGCGAAGAAACAAACATAGAGAGAGCGCCGTCAGAGAAAGCGGCTTGGAAA 940
QY      166 ProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGluLysGlnSerLeuArg 185
Db      939 CCGTACTACAAAGCTGAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 880
QY      186 AlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProValAlaProTyrAsnThr 205
Db      879 GCTTCAAGATCCGACCGCGAGATGTTCCCAAGGGCAGCGGCTCCGCTTAAACACC 820
QY      206 ThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeuLysTrpGlyLeuTyr 225
Db      819 ACCGACTTCCCTCATGATGATGATCATACAGACAGAGAGAGCGGATTTCAAAACGGCTGTAC 760
QY      226 SerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAspPheMetGluGly 245
Db      759 TTCACAGCGGCGCCCGCCCAATCCGAGACACAGCAGCATGACACTTGAAGAGAGG 700
QY      246 GlnGluGluAspGlyLysSerAspGlyMetGlyLysAspGlySerGluPheLeuGlnArg 265
Db      699 GGTGAGAGAGATGGGGGACACCGATGGGATGGAGGAGGACGAGGATTTCTGACAGCGG 640
QY      266 AspPheSerGluThrTrpGluArgTyrHisThrGluSerLeuGlnAsnMetSerLysGln 285
Db      639 GACTTCTCGGAGAGACGATGAGCGGCTACACAGAGAGAGCGCTCGAGAAATGACAGACAG 580
QY      286 GlnLeuIleLysGluTrpLeuGlnLeuGlnLysCysLeuSerArgMetGluAspGluAsn 305
Db      579 GAGCTCATCAAGAGAGTACCTGGAACCTGAGAACTGCTCTCCGGCATGGAGGAGAGAAC 520
QY      306 AsnArgLeuArgLeuGlnSerLysArgLeuGlyIAspAspAlaArgValArgGluLeu 325
Db      519 AACCGGCTCGGCTCGAGACAGCAAGCGCTGGGAGGAGCTACCGCGCTGTCCGCGAGCTG 460
QY      326 GlnLeuGlnLeuAspArgLeuArgAlaGlnAsnLeuGlnLeuThrGlnAsnGlnLeu 345
Db      459 GAGCTGAGAGCTGGAGCGCGTG-CGCCCGAGAACTTCAGTGTGACCGAGAAACGAACTG 401
QY      346 HisArgGlnGlnGlnArgAlaProLeuSerLysPheGlyAsp 359
Db      400 CACCGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 359

RESULT 11
BM474949
LOCUS      BM474949      1080 bp      mRNA      linear      EST 05-FEB-2002
DEFINITION AGENCOURT.6476734 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5562756
ACCESSION  BM474949
VERSION     BM474949.1 GI:18523991
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1080)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLMI2292 row: k column: 13  
High quality sequence stop: 404.  
Location/Qualifiers  
1..1080  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5562756"  
/clone\_lib="NIH-MGC\_88"  
/tissue\_type="Duodenal adenocarcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: small intestine; Vector: pCMV-SPORT6;  
Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally;  
oligo-dt primed. Average insert size 1,767 bp. Library  
enriched for full-length clones and constructed by Life  
Technologies. Note: this is a NIH-MGC Library."  
BASE COUNT 251 a 339 c 357 g 133 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 2,59e-71 Length: 1080  
Score: 1056.50 Matches: 222  
Percent Similarity: 88.21% Conservative: 10  
Best Local Similarity: 84.41% Mismatches: 20  
Query Match: 55.31% Indels: 13  
DB: 13 Gaps: 4  
US-09-972-758a-2 (1-359) x BM474949 (1-1080)  
QY 31 GUAATGProProGlyAlaAGluGluArgValProGluGluAspSerArgTrpGlnSerArg 50  
|||||  
Db 3 GAGCGCGCCCGCAGCGCGGAGGAGCGCGTCCCGAGAGAGACAGTGTGGCAATGAGA 62  
QY 51 AlApepProGluLeuGlyGlyArgProGlyProGluGlyGluGlySerLeuGlnSerGln 70  
|||||  
Db 63 GCGTTCCCGCCAGTTGGGTGGCGCTCGGCGCGGAGGCGGAGGCGAGCTCGAATCCCA 122  
QY 71 ProProProLeuGluThrGlnAlaCysProGluSerSerCysLeuArgGluGlyGlyLys 90  
|||||  
Db 123 CCACCTCCCTTCAGACGCCAGCGCTCTCCAGATCTAGGTGCTTGAGAGGCGCGGAG 182  
QY 91 GtGlnAsnGlyAspAspSerSerAlaGlyGlyAspPheProProAlaGluValGlu 110  
|||||  
Db 183 GGCAGAAATGGGAGCAGACTGTCGCTGGCGGCGACTTCCGCGCGCAGAGAAGTGC 242  
QY 111 ProThrProGluAlaGluLeuLeuAlaGlnProCysHisAspSerGluAlaSerLysLeu 130  
|||||  
Db 243 CCGACGCGCCGAGCGGAGCTGCTCCGCCCGCTTGTCAATGACTCCAGCGCGAGTAAGTG 302  
QY 131 GtAlaProAlaAlaGlyGlyGluGluGluGluGluGluGluGluGluGluGluGluGlu 150  
|||||  
Db 303 GGGGCTCTCGCGCAGGGGGGCGAAGAGTGGGGGACACGACGAGAGACAGCTGGGGAG 362  
QY 151 LysLysHisArgArgArgProSerLysLysLysArgHisTrpLysProTyrTyrLysLeu 170  
|||||  
Db 363 AAAAAACATAGAGAGAGCGCCGTCCAAAGAGAGCGGCGCATTTGGAACCGTACTACAGCTG 422  
QY 171 ThrTPGluGluLysLysLysPheAspGluLysGlnSerLeuArgAlaSerArgIleArg 190  
|||||

Db 423 ACCTGGAGAGAGAGAA-AAGTTGCAGCGAAAAACAGCGCTTCGAGCTTCAAGATCCGA 481  
QY 191 AlAGluMetPheAlaLysGlyGlnProValAlaProTyrAsnThrThrGlnPheLeuMet 210  
|||||  
Db 482 GCCGAGATTTTGGCCCAAGGGCCAGCCGTCGGCCCTATAC-ACACGCGAGTTCTCATG 540  
QY 211 AspAspHisAspGlnGluGluProAspLeuLysThr-GlyLeuTyrSerLysArg-Ala 230  
|||||  
Db 541 GATGATCAGCAGCCACGAGAGAGCCGATCTCAAAACCGGCGCTGTACTCCACGGGGCGG 600  
QY 230 lAlaLysSerLysAspThrSerAsp-AspAsp-PheMet---GluGluGlyGlyGlu 248  
|||||  
Db 601 CCGCCAAATCCAGCAGCACACCGAATGACGACCTTCATGCGGAGAAAAAGGGCTGGAAGA 660  
QY 248 uAspGlyGlySerAsp---GlyMetGlyLysAspGlySerGlu---PheLeuGlnArgAs 266  
|||||  
Db 661 GGATGGGGGCCACCCATGGGGATGGGAGAGGAGCGCAACAAAGTTTCTTGCGAGGGGA 720  
QY 266 pPheSerGluThrTyrGluArgTyrHisThr-----GluSerLeuGlnAsnMe 282  
|||||  
Db 721 CCTTTCGCCGAGAGAGAACCAAGGCGGAGCCACCGCGGAGAAAGACCGCGCAAAACCT 780  
QY 282 t 282  
Db 781 G 781  
RESULT 12  
Bg165450 949 bp mRNA linear EST 06-FEB-2001  
LOCUS 602345830F1 NIH-MGC\_89 Homo sapiens cDNA clone IMAGE:4456062 5',  
DEFINITION mRNA sequence.  
ACCESSION Bg165450  
VERSION Bg165450  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 949)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLMI0250 row: k column: 07  
High quality sequence stop: 622.  
Location/Qualifiers  
1..949  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4456062"  
/clone\_lib="NIH-MGC\_89"  
/tissue\_type="hypernephroma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: kidney; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dt primed.  
Average insert size 1.3 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH-MGC Library."  
BASE COUNT 243 a 257 c 335 g 114 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 2.96e-70 Length: 949  
Score: 1042.00 Matches: 239  
Percent Similarity: 89.49% Conservative: 8

Best Local Similarity: 86.59% Mismatches: 21  
 Query Match: 54.55% Indels: 13  
 DB: 12 Gaps: 1

US-09-972-758a-2 (1-359) x BG165450 (1-949)

QY 77 GlnAlaCysProGluSerSerCysLeuArgGluGluLysGluGlnAsnGlyAspAsp 96  
 |||  
 DB 3 CAGGCTGTCCAGAAATCTACTGCTGAGAGAGGCGCAGAGGCCAGAAATGGGAGCAGC 62  
 |||  
 QY 97 SerSerAlaGlyLysAspPheProProAlaGluValGluProThrProGluAlaGlu 116  
 |||  
 DB 63 TCCTCCGCTGGCGGCACCTTCCCGCGCGCGCAGAAAGTGAACCGCAGCGCCGAGGC- GAG 121  
 |||  
 QY 117 LeuLeuAlaGluProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGly 136  
 |||  
 DB 122 CTCTGTGCCCCAGGCTGTGCTGATGCTCGAGGCCAGTAAAGTG- GGGCTCTCTGCGCAGGG 180  
 |||  
 QY 137 GlyGluGluGluProGlyGlnGlnGlnArgGluLeuGlyLysLysLysHisArgArg 156  
 |||  
 DB 181 GGGGAAGAGAGAGTGGGGACAGCAGACAGACAGCTGGGGAGAAAAAACAATAGAGAGCGC 240  
 |||  
 QY 157 ProSerLysLysLysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLys 176  
 |||  
 DB 241 CCCTCCAGAGAAGAGCGGCA- TGGAAACCTTACTACAGCTGAGCTGGGGAAGAGAAA 299  
 |||  
 QY 177 LysPheAspGluLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLys 196  
 |||  
 DB 300 AAGTTGACGAGAGAAACAGAACCTTCGAGCTTCAAGATCCGAGCCGAGATGTTCCGCCAG 359  
 |||  
 QY 197 GlyGlnProValAlaProTyrAsnThrThrGlnPheLeuMetLysAspHisAspGlnGlu 216  
 |||  
 DB 360 GGCACAGCCGGTCCGGCCCTATACACACAGCAGCTTCTCATGATGATCAGCAGCAGAG 419  
 |||  
 QY 217 GluProAspLeuLysThrGlyLeuTyrSerLysArgAlaAlaLysSerAspAspThr 236  
 |||  
 DB 420 GAGCCGAGTCTCAAAACCGGCTGTACTCCAAACGGGCGC- CGCAAAATCCGACGACCC 477  
 |||  
 QY 237 SerAspAspAspPheMetGluGluGlyGlyGluGluAspGlyLysSer-AspGlyMetGlu 256  
 |||  
 DB 478 AGGATGACGACCTTCAATGAGAAAGGGGGTGAAGAGAGCTGGGGCAGCCGATGGATGGG 537  
 |||  
 QY 256 YelLysAspGlySerGluPheLeuGlnArgAspPheSer- GluThrTyrGluArgTyrHis- 275  
 |||  
 DB 538 AGGGGACGGAGGC- GAGTTTCTGACAGGGGAGCTTCTGGGACAGCTGAGACCGGTACAC 596  
 |||  
 QY 276 ThrGluSerLeuGlnAsnMetSerLys---GlnGlu-LeuIleLysGluTyrLeuGluLe 294  
 |||  
 DB 597 ACGGAGAGCCTGCGAACAATGAGCCCAAGCCGGGAAGCTCATCAAGAGACTGACTGAGCC 656  
 |||  
 QY 294 uGluLysCysLeuSerArgMetGluAsp-GluAsn-AsnArgLeuArgLeuGluSerLys 313  
 |||  
 DB 657 TGAAGAGTTGCTCTGCGCATGAGAGCCGAGAACCAACCGGCTGAGGCTGGAGAGCAG 716  
 |||  
 QY 314 ArgLeuGlyLysAspAspAla-ArgValArgGluLeuGluLeuGluLeuAspArgLeu 333  
 |||  
 DB 717 CGGCTGGGTGGCAGCAGCGCGCTGTGGGAGCTGGAAGCGGAGCCGCTGTGG 776  
 |||  
 QY 333 gAlaGluAsnLeuGlnLeuLeuThrGluAsnGlu 344  
 |||  
 DB 777 GCGCGGAGACCTTCACTGTGTGAGCGAGAACCA 810  
 |||

RESULT 13  
 AL576112/c 986 bp mRNA linear EST 16-FEB-2001  
 LOCUS AL576112 L1L\_NFL006\_PL2 Homo sapiens cDNA clone CSDD1072YL06 3  
 DEFINITION prime, mRNA sequence.

ACCESSION AL576112  
 VERSION AL576112.1 GI:12937932  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE Mammalia: Eutheria: Primates: Catarrhini; Hominoidea: Homo.  
 1 (bases 1 to 986)  
 AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES  
 source  
 1. 986  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="CSDD1072YL06"  
 /clone\_1ib="L1L\_NFL006\_PL2"  
 /issue\_type="placenta"  
 /note="Vector: pCMVSPORT 6; Site\_1: NotI; 1st strand cDNA  
 was primed with a NotI-oligo(dT) primer. Five prime end  
 enriched, double-stranded cDNA was digested with Not I and  
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6  
 vector. Library was normalized. Library was constructed by  
 Life Technologies. Contact : Feng Liang Life Technologies,  
 a division of Invitrogen 9800 Medical Center Drive  
 Rockville, Maryland 20850, USA Fax : (1) 301 610 8371  
 Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com"

BASE COUNT 216 a 284 c 217 g 265 t 4 others

ORIGIN

Alignment Scores:

Pred. No.: 2,93e-66 Length: 986  
 Score: 990.00 Matches: 197  
 Percent Similarity: 95.26% Conservative: 4  
 Best Local Similarity: 93.36% Mismatches: 8  
 Query Match: 51.83% Indels: 3  
 DB: 9 Gaps: 1

US-09-972-758a-2 (1-359) x AL576112 (1-986)

QY 149 GlyLysLysLysHisArgArgProSerLysLysLysArgHisTrpLysProTyrTyr 168  
 |||  
 DB 985 GGAAGAAAAACATRGAAACCGCGTCCA-----AGAGAAGCGGCAATTGAACCGTACTAC 932  
 |||  
 QY 169 LysLeuThrTrpGluGluLysLysLysPheAspGluLysGlnSerLeuArgAlaSerArg 188  
 |||  
 DB 931 AAGCTGACCTGGGAAAGAAAGAAAGTTGACAGAAACAGAGCCTTGAGCTTAAG 872  
 |||  
 QY 189 IleArgAlaGluMetPheAlaLysGlyGlnProValAlaProTyrAsnThrThrGlnPhe 208  
 |||  
 DB 871 ATCCGAGCGAGATGTTCCGCCAAGGGCCAGCGGTCGCCCTCATTAACACACGACGCTTC 812  
 |||  
 QY 209 LeuMetAspAspHisAspGlnGluGluProAspLeuLysThrGlyLeuTyrSerLysArg 228  
 |||  
 DB 811 CTCATGATGATATACCAACAGAGAGCGGATCTCAAAACCGGCTGTACTTCAAGCGG 752  
 |||  
 QY 229 AlaAlaAlaLysSerAspAspThrSerAspAspPheMetGluGluGlyGlyGluGlu 248  
 |||  
 DB 751 GCGCGCGCAAAATCCAGCAGCACCAGCATGAGACTTTCATGGAAGAGGGGTGAGAG 692  
 |||  
 QY 249 AspGlyLysSerAspGlyMetGlyLysAspGlySerGluPheLeuGlnArgAspPheSer 268  
 |||  
 DB 691 GATGGGGGACGATGAGATGGAGGCGCAGCGACGACGAGTTCGACGCGGACTTCTCG 632  
 |||  
 QY 269 GluThrTyrGluArgTyrHisTrpGluSerLeuGlnAsnMetSerLysGlnGluLeuIle 288  
 |||  
 DB 631 GAGACTTACAGCGGATACCAACGAGAGACCTTCACAAATATGACAGAGAGGCTCAATC 572  
 |||  
 QY 289 LysGluTyrLeuGluGluGluLysCysLeuSerArgMetGluAspGluAsnAsnArgLeu 308  
 |||  
 DB 571 AASGATACCTGGAAGCTGGAAGTCTCTCGCGCATGAGAGAGCAGAAACCGGCTG 512  
 |||  
 QY 309 ArgLeuGluSerLysArgLeuGlyLysAspAspAlaArgValArgGluLeuGluGluGlu 328  
 |||



Db		511	CCGCTGGAGACCAAGCGGCTGGCTGGCAGACACGCCGCTGTGC GGAGACTGGAGCTGGAGC	452
Oy		329	LeuAspArgLeuAraGAlaGlnAsnLeucGlnleuleuthrGluAsnGlnLeuHisArgGln	348
Db		451	CTGACACCGCTG-CGCGCGCAGAACCTCCAGCTGCTGACCGAGAACGAACTCACCGGACG	393
Oy		349	GlnGlnLuarGAlaProLeuSerLysPheGlyasp	359
Db		392	CAGAGCGCAGCGCCGCTTTCACAATTTTGAGAC	360
RESULT 14				
BMI29736				
LOCUS				
DEFINITION		BMI29736	557 bp	mRNA linear EST 12-MAR-2002
		lft2e01.y1	Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens	
		cDNA clone IMAGE:5677152 5'	similar to TR:094992 094992 HIS1	
		PROTEIN.	:	mRNA sequence.
ACCESSION		BMI29736		
VERSION		BMI29736.1	GI:17124288	
KEYWORDS		EST.		
SOURCE				
ORGANISM			human.	
		Homo sapiens		
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE				
AUTHORS				
		1 (bases 1 to 557)		
		Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Keastner,K.,		
		Lemishka,I., Scaerac,M., Brestfield,J., Gradwohl,G., Clifton,S.,		
		Hillier,L., Marra,M., Pape,D., Wyllie,T., Martin,J., Blisstein,A.,		
		Schmitt,A., Theising,B., Rittler,E., Ronko,I., Bennett,J., Cardenas		
		,M., Gibbons,M., McCann,R., Cole,R., Tsagarelisvilli,R., Williams,T.		
		, Jackson,Y. and Bowers,Y.		
TITLE				
JOURNAL		Endocrine Pancreas Consortium		
COMMENT		Unpublished (2000)		
		Other-ESTs: lft2e01.x1		
		Contact: Douglas Melton, Klaus H. Keastner, & Hiroshi Inoue		
		Harvard University, Howard Hughes Medical Institute		
		Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,		
		MA 02138		
		Tel: 617-495-1812		
		Fax: 617-495-8557		
		Email: dmelton@biohp.harvard.edu		
		Library was constructed by Dr. Douglas Melton DNA sequencing by:		
		Washington University Genome Sequencing Center for Information on		
		obtaining a clone please contact: Juliana Brown		
		(brownjef@s.harvard.edu) This sequence now available from the IMAGE		
		consortium, for clone orders contact: infoimage.llnl.gov		
		High quality sequence stop: 443.		
FEATURES				
Source				
		1..557		
		/organism="Homo sapiens"		
		/db_xref="taxon:9606"		
		/clone="IMAGE:5677152"		
		/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"		
		/sex="Both		
		/tissue_type="Islets of Langerhans"		
		/dev_stage="Adult"		
		/lab_host="DH10B"		
		/note="Organ: Pancreas; Vector: pSPORT1; Site1: Not 1;		
		Site_2: Sal I; Starting library constructed using		
		SuperScript Plasmid Library kit (Life Technologies). cDNA		
		made by oligo-dT priming. Size-selected by column		
		fractionation; average insert size 1.08 kb. Library was		
		amplified once on solid support and plasmid DNA from		
		library was prepared. The library DNA was normalized by		
		method #4 from Bonaldo, Lennon, and Soares 1996 Genome		
		Research 6:791-806; 0.5 microgram single-stranded library		
		plasmid DNA was mixed with 5 micrograms PCR product		
		representing library inserts and hybridized to an EcoT 05		
		20. Single-stranded (unhybridized) plasmids were isolated		
		by hydroxyapatite chromatography and used to make this		
		library."		
BASE COUNT		142 a	150 c	193 g 72 t
ORIGIN				

Alignment Scores:			
Pred. No.:	2.02e-64	Length:	557
Score:	963.00	Matches:	184
Percent Similarity:	99.46%	Conservative:	0
Best Local Similarity:	99.46%	Mismatches:	1
Query Match:	50.42%	Indels:	0
DB:	13	Gaps:	0
US-09-972-758A-2 (1-359) x BM129736 (1-557)			
OY	157	ProSerLysLysLysArgHisTSTPrLysProTyrTTLysLeuThrTrpGluLysLys	176
Db	3	CCGTCGAAGAAGAACGGCGCATTTGGAAACCGTACTCAACGCTGACTCGGCAAGACAGAA	62
OY	177	LysPheAspGluLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLys	196
Db	63	AAGTTTCGACGAAACACAGCCCTTCGACCTTCAACGATCCGACGCGAATGTCGCCAAG	122
OY	197	GlyGlnProValAlaProTyrAsnThrThrGlnPheLeuMetLeuAspHisAspGlnGlu	216
Db	123	GGCCAGCGCGTGGCGCCCTATTAACACCGACGATTCCTCATGGATGATCACACAGAGAG	182
OY	217	GluProAspLeuLysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspSphr	236
Db	183	GAGCCGGAAGCTCAAAACCGCGCTGTACTCCAAAGCGCGCCGCCAAATCCGACGACACC	242
OY	237	SerAspAspSphMetGluGlnGlyGlnGluLysGlyGlySerAspGlyMetGly	256
Db	243	AGCGATGACGACTTCATGTAGAAAGGGCGGTGAGGAGATGGGGGACGATGGCGATGGGA	302
OY	257	GlyAspGlySerGlnPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThr	276
Db	303	GGGACGCGCAGCAGATTTCTCCACGGGACATTCCTCGGAGCTACGACGCGTACCACAG	362
OY	277	GluSerLeuGlnAsnMetSerLysGlnGluLeuLleLysGluTyrLeuGluLeuGlnLys	296
Db	363	GAGAGCGTCGACGACATATAGCAGACGAGAGCTCATCAAGAGACTGACTGGACTGGAGAAG	422
OY	297	CysLeuSerArgMetGluAspGluAsnAsnArgLeuArgLeuGlnSerLysArgLeuGly	316
Db	423	TGCCTCTTCGCCCATATGAGAGACGAACACCGCGCTCGAGACGACGCGCTGCGT	482
OY	317	GlyAspAspAlaArgValArgGluLeuGlnLeuGlnLeuAspArgLeuArgAlaGlnAsn	336
Db	483	GGCGACGACCGCGCGGTGCGGAGCTGAGCTGAGCGACCGGCTCGCGCCGAGAAC	542
OY	337	LeuGlnLeuLeuThr 341	
Db	543	CTCCAGCTCGTCGACC 557	
RESULT 15			
LOCUS	BF167480	762 bp	mRNA linear EST 30-OCT-2000
DEFINITION	601175078F1 NCI_GAP_Lu29 Mus musculus cDNA clone IMAGE:3995683 5',		
ACCESSION	BF167480		
VERSION	BF167480.1	GI:11047832	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	1 (bases 1 to 762)		
JOURNAL	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: <a href="mailto:cga@bbs.fda.nih.gov">cga@bbs.fda.nih.gov</a>		
	Tissue Procurement: Gilbert Smith, Ph.D.		
	cDNA Library Preparation: Life Technologies, Inc.		
	GenBank library by: The I.M.A.G.E. Consortium (LLNL)		
	DNA sequencing by: Invitae Genomics, Inc.		

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLM9215 row: 1 column: 20  
High quality sequence stop: 701.  
Location/Qualifiers

## FEATURES

Source

1. 762

/organism="Mus musculus"

/strain="CZECH II"

/db\_xref="taxon:10090"

/clone="IMAGE:3995683"

/clone\_lib="NCLCGAP\_Lu29"

/tissue\_type="spontaneous tumor, metastatic to mammary."

/stem\_cell\_origin=""

/lab\_host="DH10B"

/note="Organ: Lung; Vector: pCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 200 a 208 c 248 g 106 t  
ORIGIN

## Alignment Scores:

Pred. No.: 1.57e-63 Length: 762  
Score: 953.00 Matches: 186  
Percent Similarity: 82.63% Conservative: 9  
Best Local Similarity: 78.81% Mismatches: 37  
Query Match: 49.90% Indels: 4  
DB: 12 Gaps: 3

US-09-972-758a-2 (1-359) x BE167480 (1-762)

OY 1 MetaIaGIuPProPheIuSerGIuTyGIuHISGIuProGIuInThrSerAsnCysThrGly 20  
DB 16 ATGGCGCGAGCCACTCTTGACAGAACATCAACACACGCCCTCAACTAGCACTGTACAGT 75  
OY 21 AlaIaIaIaValGIuGIuLeuAsnProGIuArgProGIuAlaGIuGIuArgVal 40  
DB 76 GCTGCTGTGTTCATGAAGACATACCTCTGAGCGCCCCCAAGCGGAGGAAGGGGTG 135  
OY 41 ProGIuGIuAspSerArgTrpGIuSerArgAlaPheProGIuLeuGIuGIuArgProGIy 60  
DB 136 CCCAAGGAGAGAGAGTGGCAATCGAGAGCTGCTTGCAGTCCGGTAGCCGTCCAGG 195  
OY 61 ProGIuGIuGIuGIuSerGIuSerGIuProProGIuGIuGIuGIuGIuGIuGIuGIuGIu 80  
DB 196 CAGAGGAGAGAGGGGCTGACAGACACGCTGCCCATGTCAGACGAATGCTGTCCA 255  
OY 81 GluSerSerCysLeuArgGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 100  
DB 256 GAATTGAGCTCTCGAGCAAGGCGGAGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 315  
OY 101 GIuAspPheProProGIuAlaGIuValGIuProThrProGIuAlaGIuLeuLeuAlaGIu 120  
DB 316 GGT---GCCTCCCGTGGCGGAGGAGAGCCGATGTCAGATCC-----CTCGTGCAG 366  
OY 121 ProCysHisAspSerGIuAlaSerLysLysLysLysLysLysLysLysLysLysLysLys 140  
DB 367 CCAGGTGATGACTCGGAGGCGCACCAAGAGAGAGGCTCTGCGCGGAGGAGGAGGAGGCA 426  
OY 141 TrpGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 160  
DB 427 TGGGAGACGACACAGACACCTGCGCAAGAAAAACATCGAGAGCGCCCTCAAGAGAG 486  
OY 161 LysArgHisTrpLysProTyTrpLysLysLysLysLysLysLysLysLysLysLysLys 180  
DB 487 AAGCGCATTTGGAAGCCCTACTACAGCTGACTGGAGAGAGAGAGAGAGAGAGAGAGAG 546  
OY 181 LysGIuSerLeuArgAlaSerArgIleArgAlaGIuMetPheAlaLysGIuGIuProVal 200  
DB 547 AAGCAGAGCCTGCGAGCTTCGGGGTTCGAGCCGAGATGTTCCCAAGGGCCAGCAGTT 606  
OY 201 AlaProTyAsnThrTrpGIuPheLeuMetAspHisAspGIuGIuGIuGIuGIuGIuGIu 220

DB 607 GCGCCCTATTAACACCGACGACTTCTCTCATGTGACACAGATCAGAGAGAGCCTGATCTC 666  
OY 221 LysThrGIuLeuTyTrpSerLysArgAlaIaIaLysSerAspPheThr 236  
DB 667 AAACCGGCGCTTTACCCCAAG---GGCAGCGCAATCGAGAGACA 711

Search completed: June 17, 2003, 12:58:57  
Job time : 3244 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 10:23:40 ; Search time 48 Seconds  
(without alignments)  
799.640 Million cell updates/sec

Title: US-09-972-758A-2  
Perfect score: 1910  
Sequence: 1 MAEPFLSEYQHQPOTSNCTG.....LTENELRHQGRAPLKRFGD 359

Scoring table: BIOSUM62  
Gapop 10.0 , Gapept 0.5

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCIT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCIT08\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1910	100.0	359	9	US-09-972-758-2
2	1910	100.0	359	10	US-09-745-763-34
3	134.5	7.0	546	9	US-09-976-740-44
4	134.5	7.0	546	12	US-10-023-529-44
5	134.5	7.0	546	12	US-10-023-523-44
6	130	6.8	611	10	US-09-216-393-81
7	128	6.7	2476	9	US-09-824-574-7
8	124	6.5	489	10	US-09-876-889-350
9	123.5	6.5	557	9	US-09-976-740-5
10	123.5	6.5	557	10	US-09-962-055-5
11	123.5	6.5	557	12	US-10-023-529-5
12	123.5	6.5	557	12	US-10-023-523-5
13	121	6.3	538	10	US-09-827-822-9
14	120	6.3	427	9	US-10-050-704-273
15	120	6.3	483	9	US-10-050-704-272
16	120	6.3	911	9	US-10-245-103-52
17	120	6.3	911	9	US-10-245-107-52
18	120	6.3	911	9	US-10-245-143-52
19	120	6.3	911	9	US-10-245-771-52

20	120	6.3	911	9	US-10-245-851-52	Sequence 52, Appl
21	120	6.3	911	9	US-10-245-883-52	Sequence 52, Appl
22	120	6.3	911	9	US-10-237-535-52	Sequence 52, Appl
23	120	6.3	911	9	US-10-238-183-52	Sequence 52, Appl
24	120	6.3	911	9	US-10-238-283-52	Sequence 52, Appl
25	120	6.3	911	9	US-10-238-370-52	Sequence 52, Appl
26	120	6.3	911	9	US-10-245-055-52	Sequence 52, Appl
27	120	6.3	911	9	US-10-245-147-52	Sequence 52, Appl
28	120	6.3	911	9	US-10-245-770-52	Sequence 52, Appl
29	120	6.3	911	9	US-10-245-739-52	Sequence 52, Appl
30	120	6.3	911	9	US-10-246-210-52	Sequence 52, Appl
31	120	6.3	911	9	US-10-239-196-52	Sequence 52, Appl
32	120	6.3	911	9	US-10-243-024-52	Sequence 52, Appl
33	120	6.3	911	9	US-10-245-409-52	Sequence 52, Appl
34	120	6.3	911	9	US-10-245-033-52	Sequence 52, Appl
35	120	6.3	911	9	US-10-245-621-52	Sequence 52, Appl
36	120	6.3	911	9	US-10-245-880-52	Sequence 52, Appl
37	120	6.3	911	9	US-10-245-095-52	Sequence 52, Appl
38	120	6.3	911	9	US-10-245-185-52	Sequence 52, Appl
39	120	6.3	911	9	US-10-245-437-52	Sequence 52, Appl
40	120	6.3	911	9	US-10-245-473-52	Sequence 52, Appl
41	120	6.3	911	9	US-10-245-770-52	Sequence 52, Appl
42	120	6.3	911	9	US-10-245-877-52	Sequence 52, Appl
43	120	6.3	911	9	US-10-246-976-52	Sequence 52, Appl
44	120	6.3	911	9	US-10-243-320-52	Sequence 52, Appl
45	120	6.3	911	9	US-10-242-743-52	Sequence 52, Appl

ALIGNMENTS

RESULT 1  
US-09-972-758-2  
; Sequence 2, Application US/09972758  
; Patent No. US20020160497A1  
; GENERAL INFORMATION:  
; APPLICANT: Case Western Reserve University  
; APPLICANT: Montano, Monica  
; APPLICANT: Wiltman, Bryan  
; TITLE OF INVENTION: Suppressors of Human Breast Cancer Cell Growth  
; FILE REFERENCE: 27708/04004  
; CURRENT APPLICATION NUMBER: US/09/972,758  
; CURRENT FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: US 60/238,187  
; PRIOR FILING DATE: 2000-10-05  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 359  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-972-758-2

Query Match	100.0%;	Score 1910;	DB 9;	Length 359;
Best Local Similarity	100.0%;	Pred. No. 1.3e-122;		
Matches	359;	Conservative 0;	Mismatches 0;	Gaps 0;
Db	1	MAEPFLSEYQHQPOTSNCTGAAVQEEINPRPGAERPEEDSRMOSRAFPOLGPRG	60	
Qy	1	MAEPFLSEYQHQPOTSNCTGAAVQEEINPRPGAERPEEDSRMOSRAFPOLGPRG	60	
Db	1	MAEPFLSEYQHQPOTSNCTGAAVQEEINPRPGAERPEEDSRMOSRAFPOLGPRG	60	
Qy	61	PEGGSLESPQPPLOTACPESSCLRGGEKQNDSSAGDPPPAVEPTPAELLAQ	120	
Db	61	PEGGSLESPQPPLOTACPESSCLRGGEKQNDSSAGDPPPAVEPTPAELLAQ	120	
Qy	121	PCHDSEASKIGAPPAAGEEEMGQOOLGKKRRRSRKKRHHKPYKLTWEKKKFE	180	
Db	121	PCHDSEASKIGAPPAAGEEEMGQOOLGKKRRRSRKKRHHKPYKLTWEKKKFE	180	
Qy	181	KOSLRASRIRAEKAPAPVNTTQFLMDHDHQQEPPDLKTGLYSKRAAKSDTSDDD	240	
Db	181	KOSLRASRIRAEKAPAPVNTTQFLMDHDHQQEPPDLKTGLYSKRAAKSDTSDDD	240	

QY 241 FMEEGEDGSDGCGDSEFLORDFSTYRYHTESLONNSKOEILKEYLELEKCLSR 300  
DB 241 FMEEGEDGSDGCGDSEFLORDFSTYRYHTESLONNSKOEILKEYLELEKCLSR 300  
QY 301 MEDENNRLRLLESKRGLGDDARVRELELDRLRAENLQLTENELHROOERAPLSKFGD 359  
DB 301 MEDENNRLRLLESKRGLGDDARVRELELDRLRAENLQLTENELHROOERAPLSKFGD 359

## RESULT 2

US-09-745-763-34  
Sequence 34, Application US/09745763  
Patent No. US20020065394A1

## GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth  
McCoy, John M.  
Lavallee, Edward R.  
Collins-Racie, Lisa A.  
Evans, Cheryl  
Merberg, David  
Treacy, Maurice  
Spaulding, Vikki  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
ENCODING THEM  
NUMBER OF SEQUENCES: 219  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: U.S.A.  
ZIP: 02140

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/745,763  
FILING DATE: 18-Jun-2000  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Sprunger, Suzanne A.  
REGISTRATION NUMBER: 41,323  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8284  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 359 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 34:

US-09-745-763-34

Query Match 100.0%; Score 1910; DB 10; Length 359;  
Best Local Similarity 100.0%; Pred. No. 1.3e-122;  
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEPFLSYOHOPQTSNCTGAAYOEELNPERPRAEERVPEEDSRWQSRAPFOLGGRPG 60  
DB 1 MAEPFLSYOHOPQTSNCTGAAYOEELNPERPRAEERVPEEDSRWQSRAPFOLGGRPG 60  
QY 61 PEGGSELESOPPLQTOACPESSCLREGEKONGDSDSAGDPPPPAVEPTPEALLAQ 120  
DB 61 PEGGSELESOPPLQTOACPESSCLREGEKONGDSDSAGDPPPPAVEPTPEALLAQ 120  
QY 121 PCHDSEASKLGAPAGGEEWGOOOLGKKRRRRPSSKKRRHKPYKLTWEKKKFE 180  
DB 121 PCHDSEASKLGAPAGGEEWGOOOLGKKRRRRPSSKKRRHKPYKLTWEKKKFE 180

QY 181 KOSLRASRIRAEFAKGPVAPYNTQFLMDHDOEPDLKTLGYSKRAAKSDPTSD 240  
DB 181 KOSLRASRIRAEFAKGPVAPYNTQFLMDHDOEPDLKTLGYSKRAAKSDPTSD 240  
QY 241 FMEEGEDGSDGCGDSEFLORDFSTYRYHTESLONNSKOEILKEYLELEKCLSR 300  
DB 241 FMEEGEDGSDGCGDSEFLORDFSTYRYHTESLONNSKOEILKEYLELEKCLSR 300  
QY 301 MEDENNRLRLLESKRGLGDDARVRELELDRLRAENLQLTENELHROOERAPLSKFGD 359  
DB 301 MEDENNRLRLLESKRGLGDDARVRELELDRLRAENLQLTENELHROOERAPLSKFGD 359

## RESULT 3

US-09-976-740-44  
Sequence 44, Application US/09976740  
Publication No. US20020194633A1

## GENERAL INFORMATION:

APPLICANT: Lees, Ann M.  
Lees, Robert S.  
APPLICANT: Law, Simon W.  
APPLICANT: Arjona, Aubal A.

TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING  
PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING  
ATHEROSCLEROSIS  
FILE REFERENCE: 10797-004001  
CURRENT APPLICATION NUMBER: US/09/976,740  
CURRENT FILING DATE: 2001-10-12  
PRIOR APPLICATION NUMBER: 09/616,289  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: US 08/979,608  
PRIOR FILING DATE: 1997-11-26  
PRIOR APPLICATION NUMBER: US 60/031,930  
PRIOR FILING DATE: 1996-11-27  
PRIOR APPLICATION NUMBER: US 60/048,547  
PRIOR FILING DATE: 1997-06-03  
NUMBER OF SEQ ID NOS: 53  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO: 44  
LENGTH: 546  
TYPE: PRT  
ORGANISM: Homo sapiens

US-09-976-740-44

Query Match 7.0%; Score 134.5; DB 9; Length 546;  
Best Local Similarity 21.9%; Pred. No. 0.15;  
Matches 86; Conservative 51; Mismatches 134; Indels 121; Gaps 14;

QY 20 GAAVQELNPERPPGAERVPEDSRWQSRAPFOLGGRPGEGSLESQPPLOTQAC 79  
DB 9 GAAVQELNPERPPGAERVPEDSRWQSRAPFOLGGRPGEGSLESQPPLOTQAC 79  
QY 80 PESSCLREGEKONGD-----DSSAGDPPPPAVEPTPEALLAQ 121  
DB 55 PEGQAARTAGSALRDVSEELSRQLEDLSTYVCVNNQGGEGDGAQEP-----AEP 107  
QY 122 CHDSASKL-----GAPAGGEEWGOO-----OOOROLGKKRRRRPSSKKRRH 163  
DB 108 -EDAEKSTVYARNGEPPTPVVNGEKPSGDPNTEIRSDSDVGRDRRPPQEKKAK 166  
QY 164 W-----KPYKLTWEKKKFEKOSLRASIRAEFAKGPVAPYNTQ 207  
DB 167 GLGKEITLIMQTLNTLSTPEERLALCKYALLLEHNRNSQKMLLOKKQ-----SQ 219  
QY 208 FLMDHDOEPDLKTLGYSKRAAKSDPTSDDEMEGEGEDGSDGCGDSEFLORDF 267  
DB 220 LV-----QEKDHLRGEHSRAVLARSK-----LESLREL 248  
QY 268 SETYRYHTESLONNSKOEILKEYLELEKCLSRMEDENNRLRLLESKRGLGDDARVRELE 327  
DB 249 QR-----HNRLKKEGVOARAEEFEKREKVTSHQVLTNDIOLQLOMEQHNRNSKLRQENM 303  
QY 328 ELDRRAENLQLTENELHROOERAPLSKFGD 359



CURRENT FILING DATE: 1998-12-18  
EARLIER APPLICATION NUMBER: 08/394,825  
EARLIER FILING DATE: 1997-12-19  
NUMBER OF SEQ ID NOS: 364  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 81  
LENGTH: 611  
TYPE: PRT  
ORGANISM: Toxoplasma gondii  
US-09-216-393-81

Query Match 6.8%; Score 130; DB 10; Length 611;  
Best Local Similarity 22.2%; Pred. No. 0.34;  
Matches 82; Conservative 58; Mismatches 155; Indels 74; Gaps 15;

QY 13 POTSCTGAAYQVELN-PEPPGAEEVP-----EDSRMOSRAF-OLGGRPE 62  
DB 189 PRFSKDYCCFQAAHLSLPGSLSSPSSPISVTMDYALFDSASPLHAGELSLPGAV 248  
QY 63 GEGSLESQP---PLQTQACPESSCLREGKQNGDDSSAGGDPPEAVEPEPEAEIL 118  
DB 249 SASERLLTAPAEIGSASACLSVSC-----GFGEMSPITADTTR- 287  
QY 119 AOPCHDSASKLGAPAAAGC---EWMGOQROLCKKRRRPPSKKKHMKPYKLTWEE 174  
DB 288 ---HDAEERERRRAEEERERERERERERERERERERERERERERERERERER 339  
QY 175 K---KKPFKOSLRA---SRIABEMFAKGPVAPYNTOTFLMDHDOEEDPLKGLYSK 227  
DB 340 KAROEDEDERERRRVEEKAROREDEDERRRRVEEKARQ---KEEERERRRVEEK 395  
QY 228 RAAAKSDTSDDEME---GGEFDGSDGNGGSDGSEFLORFSETYE--RYHTESLON 281  
DB 396 AROREERERRRVEEKAROREDEDERRRRVEEKAROREDEDERRRRVEEKARQ 455  
QY 282 MSKOELIKEYLELEKCLSMEDENRRLLESKRLGGDARVLELELDRLAENLQJLT 341  
DB 456 REEERERRRVEEKERERERERERERERERERERERERERERERERERER 505  
QY 342 ENELHROE 350  
DB 506 EKEREROE 514

RESULT 7  
US-09-824-574-7  
Sequence 7, Application US/09824574  
Publication No. US20030077800A1

GENERAL INFORMATION:  
APPLICANT: Rouleau, Natalie  
APPLICANT: Mollanen, Anu-Maarit  
APPLICANT: Palvimo, Jorma J.  
APPLICANT: Jone, Olli A.  
TITLE OF INVENTION: ARIPI Gene and Protein  
FILE REFERENCE: 2630-109  
CURRENT APPLICATION NUMBER: US/09/824,574  
CURRENT FILING DATE: 2001-04-03  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 7  
LENGTH: 2476  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-824-574-7

Query Match 6.7%; Score 128; DB 9; Length 2476;

Best Local Similarity 19.7%; Pred. No. 2.2;  
Matches 78; Conservative 52; Mismatches 130; Indels 136; Gaps 13;

QY 43 EDSRMSRAFPLGGRPEGE---GSLSQPPLQTQACPESSCLREGKQNGDDSSA 99  
DB 1142 EDNRKOKORTSARKKTKGTNTEKKRNSLRATPKRQOVITSSSSDGGDDONSAGESSD 1201

QY 100 GGDPPP-----PAVEP-----TPE-----AELIA 119  
DB 1202 EQKIKPYENVLPSHTGFCOSSGDEALSKSVPATVDDDDDDNNENNIAKKMLEELKA 1261  
QY 120 QPCHDSASKLGAPAGEEEMGOQ-----ROLGKKKRRRPPSKK- 161  
DB 1262 NLSSDEGSSDDEPPGGKKRIRIGQSESPADDELREQLAVQNVNSESSESSKPK 1321  
QY 162 --RHMKPYKLTWEEKKFDEKQSLRASRIABEMFAKGPVAPYNTOTFLMDHDOEEDPD 219  
DB 1322 RYRHLRLKHLLTSDGEGEK---PTKREKHEAKR----- 1356  
QY 220 LKTLGYSKRAAKSDTSDDEPMEGGEDGSDMGGSDGSEFLORFSETYERYHTESL 279  
DB 1357 ---NRRKVSSEDSEDTDFQESGVSEVSE-----SEDDQRPRTSAKKALEEN 1402  
QY 280 QMSKOELIKEYLELEKLSR-----MEENRRLLESKRL 315  
DB 1403 QRSYKOKKRRRIKQVEDSSSEKNSHSEEDKESGDEDEDEDEDEDEDEDEDEDEDE 1458  
QY 316 GGDARVLELELDRLAENLQJLTENELHROER 351  
DB 1459 GKGRKIKIKI-LKDKLTFE-----TONALKEEER 1488

RESULT 8  
US-09-876-889-350

Sequence 350, Application US/09876889  
Patent No. US20020076715A1

GENERAL INFORMATION:  
APPLICANT: Benson, Darin R.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: King, Gordon E.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN  
FILE REFERENCE: 210121.466C3  
CURRENT APPLICATION NUMBER: US/09/876,889  
CURRENT FILING DATE: 2001-06-06  
NUMBER OF SEQ ID NOS: 353  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 350  
LENGTH: 489  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-876-889-350

Query Match 6.5%; Score 124; DB 10; Length 489;  
Best Local Similarity 22.7%; Pred. No. 0.68;  
Matches 71; Conservative 52; Mismatches 146; Indels 44; Gaps 12;

QY 7 SEYOHQPOTSCTGAAYQEE-----LNPER---PGAEERPE-EDSRMOSRAF 52  
DB 3 SEYISGDO-SPDGCATPYQDERDSDGSDGEDVYNQSHSSDPTGSYERHSSENTSDSEBEL 61  
QY 53 POLGGRPEGEGSLESQPPPLQTQACPESSCLREGKQNGDDSSAGGDPPEAVEPEPT 112  
DB 62 PK-----PRISDS-ESFDDP-RNOASDSN--ELPKPRVSDSESGQKGPASDSPT 111  
QY 113 PEAEILQOPCHDSASKLGAPAGEEEMGO-----QOROLGKKKRRRPPSKKKRRHKPY 167  
DB 112 DASRHKQKPESSDDSDRENK---GEDTEMONDSFHSNDRRKFFHSDSSEEEHKKOK 167  
QY 168 YKLTWEEKKKPFDEKQSLRASRIABEMFAKGPVAPYNTOTFLMDHDOEEDPLKGLYSK 227  
DB 168 MDSDEDEKEGEFEKAVAKKAVALSDSEDEKASAKSKRVVSDADSDSDAVSKGREK 227  
QY 228 RAAAKSDTSDDEPMEGGEDG---GSDMGGSDGSEFLORF-----SEYERYHTE 277  
DB 228 TIASDSEDEACKELSDKKNEKDLFGSDSESGNEENLADIFGSDGDEEEETFTGNQE 287  
QY 278 SLQNSKOELIKE 290

```

Db      288 DLEEKGETGYKE 300

RESULT 9
US-09-976-740-5
: Sequence 5, Application US/09976740
: Publication NO. US20020194633A1
: GENERAL INFORMATION:
:   APPLICANT: Lees, Ann M.
:   APPLICANT: Lees, Robert S.
:   APPLICANT: Law, Simon W.
:   APPLICANT: Arjona, Anibal A.
: TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
: TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
: FILE REFERENCE: 10797-004001
: CURRENT APPLICATION NUMBER: US/09/976,740
: CURRENT FILING DATE: 2001-10-12
: PRIOR APPLICATION NUMBER: 09/616,289
: PRIOR FILING DATE: 2000-07-14
: PRIOR APPLICATION NUMBER: US 08/979,608
: PRIOR FILING DATE: 1997-11-26
: PRIOR APPLICATION NUMBER: US 60/031,930
: PRIOR FILING DATE: 1996-11-27
: PRIOR APPLICATION NUMBER: US 60/048,547
: PRIOR FILING DATE: 1997-06-03
: NUMBER OF SEQ ID NOS: 53
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 5
: LENGTH: 557
: TYPE: PRT
: ORGANISM: Oryctolagus cuniculus
US-09-976-740-5

Query Match      6 5%: Score 123.5; DB 9; Length 557;
Best local Similarity 22.6%: Pred. NO. 0.85;
Matches 91: Conservative 53; Mismatches 123; Indels 135; Gaps 20;

QY 20 GAAVVOELNPERPPGAEEVPE-----EDSRMOSRAFPOLGPRPEGESELSQ 70
    |||  :: |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 9 GAA---KQPNKSSFGQPEAGAEQAQGRGPARAREAGSSQAPGR--PEGQAQAKTAQ 63
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 71 PPPL-----QTQACPESSCKREGCEKGCGDSSAGDPPPAVEPTPEALLAQP 121
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 64 PGALCDVSEELROLELILSTYCV--DNNQCAPGEDGVG-----EP-PEPE----- 107
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 122 CHDSEASKL-----GAPAAAGGEEMWQOQ-----KQLGKKKRRRPSKKR 162
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 108 --DAEKSRAYVARGEPERPGTPVYVNGEKETSAPGEETILRTSDVGQRDRRRDPEKKA 165
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 163 HW-----KPYVTLTWEEKKKFKDEKOSLRASRTAEKMPAKGQPARPYNTT 206
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 166 KGLCKEITLLMOTLINTLSTPEEKLAALCKKAEILLHEHRNSQOKWKLQKKO-----S 218
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 207 QFLMDHDQEPRLDTGLYSKRAAKSDDT-----DDDFMEEGGEEDGSDGMGD 258
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 219 QLV-----QEKDHLRGHSHKAILLARSKLESCLRELQRHNNSLKEEG----- 259
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 259 GSEFLQDRFSTTYRYRTYESLQNNKSKOBLIKYLELEKCLSRMEDENNRLLESKRIGCD 318
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 260 ---VQARREBEERKEEYTSHFQMTLNDI---QLQME---QHNERNSKLROENMELA-- 306
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 319 DARVRELELELDRLRAENL-----QLLTENELEHROOE 350
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 307 -ERLKL-IEQYELREEHIDKVFHKHKLQOLQVDAKLOQAOE 346
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

```

1 Lees, Robert S.
2
3 Law, Simon W.
4 Arjona, Anibal A.
5
6 TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
7 BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
8 TREATING ATHEROSCLEROSIS
9
10 NUMBER OF SEQUENCES: 42
11
12 CORRESPONDENCE ADDRESS:
13 ADDRESSEE: Fish & Richardson P.C.
14 STREET: 225 Franklin Street
15 CITY: Boston
16
17 STATE: MA
18
19 COUNTRY: USA
20 ZIP: 02110-2804
21
22 COMPUTER READABLE FORM:
23 MEDIUM TYPE: Diskette
24 COMPUTER: IBM Compatible
25 OPERATING SYSTEM: DOS
26
27 SOFTWARE: FASTSEQ for windows Version 2.0
28
29 CURRENT APPLICATION DATA:
30 APPLICATION NUMBER: US/09/962,055
31 FILING DATE: 24-Sep-2001
32
33 PRIOR APPLICATION DATA:
34 APPLICATION NUMBER: 08/979,608
35 FILING DATE: 26-NOV-1997
36 APPLICATION NUMBER: US 60/031,930
37 FILING DATE: 27-NOV-1996
38
39 ATTORNEY/AGENT INFORMATION:
40 NAME: Myers, Louis
41
42 REGISTRATION NUMBER: 35,965
43 REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
44
45 TELECOMMUNICATION INFORMATION:
46 TELEPHONE: 617/542-5070
47 TELEFAX: 617/542-8906
48
49 INFORMATION FOR SEQ ID NO: 5:
50
51 SEQUENCE CHARACTERISTICS:
52 LENGTH: 557 amino acids
53 TYPE: amino acid
54 TOPOLOGY: linear
55 MOLECULE TYPE: protein
56 FRAGMENT TYPE: internal
57
58 SEQUENCE DESCRIPTION: SEQ ID NO: 5:
59
60 US-09-962-055-5
61
62 Query Match 6.5%; Score 123.5; DB 10; Length 557;
63 Best Local Similarity 22.6%; Pred. No. 0.85;
64 Matches 91; Conservative 53; Mismatches 123; Indels 135; Gaps 20;
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
9
```





```

:
: CURRENT FILING DATE: 2001-04-06
: PRIOR APPLICATION NUMBER: US/09/370,368
: PRIOR FILING DATE: 1999-08-09
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ ID NO 9
: LENGTH: 538
:
: TYPE: prt
:
: ORGANISM: Moloney Murine Leukemia Virus
: US-09-827-822-9

```

Query Match	6.38;	DB	10;	Length	538;
Best Local Similarity	21.38;	Pred. No.	1.2;		
Matches	86;	Conservative	49;	Mismatches	147;
				Indels	122;
				Gaps	19

```

OY 30 PERPPAEERVEEEDSRMOSRAFPOLGAPPE-----CGGS-----LESOPPLTQACPE 81
      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 117 PLEPPLS-----TPPOSSLYPALT-PSUGCAKPKQVULSDSGCLIDLITLEDPPPYRDP PPPP 172
      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 82 SSCLEGEKGONDDSSAGDPPPAVEPPPEAE LL-----AOPCHSEASKLAPAGA G 137
      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 173 SD--RDGDSG-----EATPAGEAP-----DPSMASRLGRREPPVADSTTSQAPPLTTGG 221
      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 138 EEEWGOOQROLGKKKKRRRPPSKKKKKHMKPKYKLLWEEKKCKDEKOSLRASIRAE MF AKG 197
      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 222 N-----GOLQWPFSSSDLYNMK-----NNNPSEDPGKLTALITESVLT-- 261
      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 198 QPVPYNTTQFLMDDHDOFEEDPLKTG-----LYSKRAAASDD----- 235
      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 262 -----THOPFWDDCOQLLTGLTGLLEEKORYLLERKAAVAGDDGDPRTLPVNEVDAAFP L 314
      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 236 -TSDDDFMEEGEE-----DGGSGMGWGDSFELORDFS 268
      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 315 ERDMEYTTQAGCNHLVHROLLAGLONAGSPTNLAKVYGITQGRNESFATLER-LK 373
      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 269 ETVYERYHTESLQ-----NMSKOELDIKEYLELEKLSMEDENNRL-----RLESK 313
      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 374 EAVRYRTYPDPEDEPQGETWMSMSFIQASAPDIGKLETERLEDLRKKTIGDLVREAREIFNK 433
      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 314 RLGGD--DARV--RELELELDRLAENLDOLLTENELHROQEFAPL 354
      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 434 REMPEREREIRIREKEKEERRRTTDEDEKKEKRDORRRHRESRL 477
      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

```

```

RESULT 14
US-10-050-704-273
? Sequence 273, Application US/10050704
? Publication No. US20030050442A1
? GENERAL INFORMATION:
? APPLICANT: Ruben et al.
? TITLE OF INVENTION: 62 Human Secreted Proteins
? FILE REFERENCE: P2039P1
? CURRENT APPLICATION NUMBER: US/10/050.704
? CURRENT FILING DATE: 2002-01-18
? PRIOR APPLICATION NUMBER: 09/664,524
? PRIOR FILING DATE: 2000-10-10
? PRIOR APPLICATION NUMBER: PCT/US00/08979
? PRIOR FILING DATE: 2000-04-06
? PRIOR APPLICATION NUMBER: 60/128,693
? PRIOR FILING DATE: 1999-04-09
? PRIOR APPLICATION NUMBER: 60/130,991
? PRIOR FILING DATE: 1999-04-26
? NUMBER OF SEQ ID NOS: 344
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 273
? LENGTH: 427
? TYPE: PRT
? ORGANISM: Homo sapiens
? US-10-050-704-273

```

Query Match	6.3%	Score 120	DB 9	Length 427
Best Local Similarity	24.4%	Pred. No. 1.1		
Matches 98	Conservative 35	Mismatches 141	Indels 128	Gaps 22

```

QY      0 LSYXOH-----PQR--SNCGAAAYOEL--NPERPCAERVAEEDSRMSQAPOLG 57
Db      72 LLEFETQSNVPPGFSEEGKALEEEDKEEEDKEEEDKEEEDKEEEDKEEEDKEE 128
QY      58 RPEPEEGSLTESOP-----PLIOTQACPESSCLRECEGCGNGDSSAGGDFPPAE 108
Db      129 SPCP--EASLPTEPAAQESLSQAARAVLQPGASPLPDGSEASRRPRRHG-----PPTE 182
QY      109 VEDTPAEALLAOPCHDS--EASKLGAPAGGE-----EEMGQOOROLGKKKRRRP 157
Db      183 TLPTPRENLASPSPTLVEAREVEAGTGPCLSGPGESEETGSSGCAPLLPTATRAP 242
QY      158 SKKKRRMKRYKLLTWEKKKFPDEKOSLASRSIRAEEMFAKOOPVAPYNTTQFLMDHQEE 217
Db      243 EGTRELEAP-----SEDSGSR-----TAPAGTSV-----QAO 269
QY      218 PDLCTGLYSKRAAKSDTSD--DDFMEECG--EEDCG--SDGCGD--GSEFL 263
Db      270 PVLPTDSASRGVAVVPASGDCVPSPCHNGCTCLEEGGVACLCPLPGVGGLCDVGLRFC 329
QY      264 Q-----RDFS-----ETERY--HTESLONNSKOELLKEYLELEKCLSRM 301
Db      330 NPGMDAFQGACYKHFSTRSRMEAEATQCRMYGAHLASISTREDDFL----- 376
QY      302 EDENNRLLSKRKLGDDARVALELELDRLRAEMQLLTEN 343
Db      377 ---NNRYR--EYOWIGLNDRTI-----EDGFLMSDQVPLLYTEN 409

```

```

RESULT 15
US-10-050-704-272
: Sequence 272, Application US/10050704
: Publication No. US20030050442A1
: GENERAL INFORMATION:
: APPLICANT: Ruben et al.
: TITLE OF INVENTION: 62 Human Secreted Proteins
: FILE REFERENCE: P2039P1
: CURRENT APPLICATION NUMBER: US/10/050.704
: PENDING FILING DATE: 2002-01-18
: PRIOR APPLICATION NUMBER: 09/684,524
: PRIOR FILING DATE: 2000-10-10
: PRIOR APPLICATION NUMBER: PCT/US00/08979
: PRIOR FILING DATE: 2000-04-06
: PRIOR APPLICATION NUMBER: 60/128,693
: PRIOR FILING DATE: 1999-04-09
: PRIOR APPLICATION NUMBER: 60/130,991
: PRIOR FILING DATE: 1999-04-26
: NUMBER OF SEQ ID NOS: 344
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 272
: LENGTH: 483
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-050-704-272

```

Query Match	6.3%	Score 120;	DB 9;	Length 483;
Best Local Similarly	24.4%	Pred. 0.12;		
Matches 98;	Conservative 35;	Mismatches 141;	Indels 128;	Gaps 22

QY 0 LSYQHQ-----PQT--SNCTGAAGVQEL--NPERPGAERPRVEDDSFMQSRAPQJLG 57

Db 1 LLEFETQSWMPPIPGFSEEBGKALFEEEYVEDEEEKEEEEEVEDEAL--AWPSELS 180

QY 58 RPCEGEGSLSP-----PPLQYQAPESCUREGKQNGDDSSAGDPPPAE 108

Db 124 LLEFETQSWMPPIPGFSEEBGKALFEEEYVEDEEEKEEEEEVEDEAL--AWPSELS 180

QY 181 SPSP--EASLPPEAAQEEESQAPARAVLQPGASPLPDGGEASRPVHG---PPTY 234

Db 109 VEPTEPAELLAPCHDS--EASKLAPAGE-----EAGQGOQRLGKKHRRP 157

QY 235 TLPTPERNLASTFSSTLYAEAREVEAATGPELSCVPAGESEETGSSGAGSLLPATRAP 294

Db 158 SKKRHRMKPYKLTVEEKKPFDEKQSLRASRIIAEMFAKQGVAPVNTTQFLMDHDOEF 217

```
Db      295 EGTRLEAP-----SEDNSGR-----TAPAGTSV-----OAQ 321
OY      218 PDLKTGLYSKRAAKSDPTSD--DDFMEEGG---EEDGG-----SDGMGGD-----GSFFL 263
Db      322 PVLPTDSASRGVAVVPASGDCVPSPCHNGGTCTLEEEBEGVRCCLPGYGGDLCDVGLNFC 381
OY      264 Q-----RDFS-----ETERY---HTESLQNMKSQELIKEYLELEKCLSRM 301
Db      382 NPGMDAFOGACYKHFSTRSWEAEFTQCRMYGAHLASISTPEODFI----- 428
OY      302 EDENNRRLRESKRIGGDDARVRELELDRLRAENLQLTEN 343
Db      429 ---NNRYR-EYOWIGLNDRTI---ESDFLMSDGVPLLYEN 461
```

Search completed: June 17, 2003, 10:33:41  
Job time : 50 secs



Db 629 KEEHOVMAKARQYAEERKORAEI 654

# RESULT 2

S32103

filensin - bovine

N:Alternate names: intermediate filament protein

C:Species: Bos primigenius taurus (cattle)

C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #ext\_change 10-Sep-1997

C:Accession: S32103; A40690

R:Gounari, F.; Merdes, A.; Quinlan, R.; Hess, J.; Fitzgerald, P.G.; Ouzounis, C.; George

submitted to the EMBL Data Library, March 1993

A:Description: Bovine filensin possesses primary and secondary structure similarity to

A:Reference number: S32103

A:Accession: S32103

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-755 <GOU>

A:Cross-references: EMBL:X72388; NID:g287751; PID:g287752

R:Gounari, F.; Merdes, A.; Quinlan, R.; Hess, J.; Fitzgerald, P.G.; Ouzounis, C.A.; Geor

J. Cell Biol. 121, 847-853, 1993

A:Title: Bovine filensin possesses primary and secondary structure similarity to interme

A:Reference number: A40690; MUID:93260017; PMID:8491777

A:Accession: A40690

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-622, RP, 625-755 <GOU>

A:Experimental source: lens

A:Note: sequence extracted from NCBI backbone (NCBI:132495, NCBI:132499)

C:Keywords: membrane-associated protein

Query Match 7.3%; Score 139; DB 2; Length 755;

Best Local Similarity 22.6%; Pred. No. 0.52;

Matches 88; Conservative 32; Mismatches 117; Indels 152; Gaps 16;

Db 26 EELNPRPGAEERVPEEDSRMOSRAFPOLGAPRPGEGSGLESOP-PLQTOA----- 78

414 EASPTQOGAEDVP--DGRKISKAFELKMKIKVKGPKPEPPADLYTKGRVYVS 471

79 -----CPSSCLFREGKEGONGDSSAGDPPEPAVEPTP-----EAFLLAO 120

472 GGGSEVDPGFCVFSVPKAGGVVSKGDDS-----VPPDSGVPSPOQPEPPLLEGQPPQ 526

121 PCHDSEASKIGAPAAAG-----EEWG 142

527 EKEDGLKEGCGPEPGKGGDSVKEGGPPEGKGGGVKEGGPPEGKGQVKEEG 586

143 QOO-RQLGKKHRRRRSKKRRHKKPYKLTWEKKKFFDKOSLRASRIRAEFAKQPYA 201

587 PEGKGDGVKKEGPEGEGELK-----EEGFLQEKED-----GQSP 626

202 PYNTQFLMDHDOEPDLKTGLYSKRAAKSDTSDDFMEG-----GEED-- 249

627 PHPA-----DKQDEKAKELKGLQKQ-----DQKEEGAGPCPMPVAPGEGS 671

250 -----GSDMGSGDSEFLOR-----DFSEYERHTESLQNNSKQPL 287

672 TPSSGQPOVILGSEHGARGSGSRLASPPRIAYEKVEMSIKFTSESIOYEETAV 731

288 IKEYLELEKCLSRMEDENNRLRLESKRLG 316

732 IVETM-IEK-----TRANKKRLG 748

Db

RESULT 3

I61106

involucrin - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 02-Aug-1996 #sequence\_revision 23-Aug-1996 #ext\_change 22-Jun-1999

C:Accession: I61106

R:Djian, P.; Phillips, M.; Esley, K.; Huang, E.; Simon, M.; Rice, R.H.; Green, H.

Mol. Biol. Evol. 10, 1136-1149, 1993

Query Match 7.0%; Score 134; DB 1; Length 568;

Best Local Similarity 20.9%; Pred. No. 0.76;

Matches 93; Conservative 66; Mismatches 169; Indels 118; Gaps 18;

Db 3 EPLSEYCHOPQTSNCTGAAYQEE-----LNPERPGAEERVPED 44

127 EQLRHQHQPOQESQOGLCLQOQDVLAPQELHMQHOKELQEBELPLGQOQKTPQEQ 186

45 -----SRMOSRAFPOLGGRPE-----GEGSLESOPPLQTOAC-----PESSCLR 86

187 ELLGKQKQLHVERHNPQRODELHNGOKQOQPOQEBELQVQHQKQHPE-LCLR 245

87 ESEKQNGDSSAGDPPEPAVEP-----TPAEALLQPCDSEASKLGAP 133

246 KQOQOESHERELHLGKQOQOESHERELHLGKQOQOESHERELHLGKQOQOESCEPEL 302

134 AAGSEEMQOQORO-----LGKKHNR-----RPSKKRHHKPYKLTWEKKKRP 178

303 HLGEOQHODQOQOESCEPELHLGKQOQOESCEPELHLGKQOQOESCEPELHLGKQOQOESCEPEL 362

179 -----DEKQSLRASRIRAE-----FAKQAPAPYNTQFLMDHDOEP 218

363 DEPLHNGKQOQOESCEPELQVKEKQESPEDELHNGKQOQOESCEPELQVKEKQESPEDELHNGKQOQOESCEPEL 422

219 DKTGLYSKRAAKSDTSDDFMEGEGEDGSGDGGSEFLQRFSEYERHTES 278

423 ELHLG-----KQESHERPTEDEKQESLYEPELHLGKQOQOESCEPELQVKEKQESPEDELHNGKQOQOESCEPEL 472

279 LQNSKQOELIKYLELEKLSRMEDENNRL--RLESKRLGSDAVRELELRLRAEN 336

473 LMQLKQEKASRGQELDD--SHLEQEKELLDRLQDELVNKDE-----QLERKKHKLLEN 524

337 L-----QLTENEHLRQOERAPL 354

525 LQKQEKQIKQLVPSPD--RVQETQPI 548

A:Title: The involucrin genes of the mouse and the rat: study of their shared repeats

A:Reference number: A49377; MUID:94104476; PMID:8277848

A:Accession: I61106

A>Status: translation not shown; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-568 <RES>

A:Cross-references: GB:I28818; NID:9454427; PIDN:AAA1445.1; PID:9454428

C:Comment: During the terminal differentiation of keratinocytes, this protein from th

linked envelope under the plasma membrane.

C:Superfamily: involucrin

C:Keywords: cornified cell envelope; duplication; epidermis; tandem repeat

Query Match 7.0%; Score 134; DB 1; Length 568;

Best Local Similarity 20.9%; Pred. No. 0.76;

Matches 93; Conservative 66; Mismatches 169; Indels 118; Gaps 18;

Db 3 EPLSEYCHOPQTSNCTGAAYQEE-----LNPERPGAEERVPED 44

127 EQLRHQHQPOQESQOGLCLQOQDVLAPQELHMQHOKELQEBELPLGQOQKTPQEQ 186

45 -----SRMOSRAFPOLGGRPE-----GEGSLESOPPLQTOAC-----PESSCLR 86

187 ELLGKQKQLHVERHNPQRODELHNGOKQOQPOQEBELQVQHQKQHPE-LCLR 245

87 ESEKQNGDSSAGDPPEPAVEP-----TPAEALLQPCDSEASKLGAP 133

246 KQOQOESHERELHLGKQOQOESHERELHLGKQOQOESHERELHLGKQOQOESCEPEL 302

134 AAGSEEMQOQORO-----LGKKHNR-----RPSKKRHHKPYKLTWEKKKRP 178

303 HLGEOQHODQOQOESCEPELHLGKQOQOESCEPELHLGKQOQOESCEPELHLGKQOQOESCEPEL 362

179 -----DEKQSLRASRIRAE-----FAKQAPAPYNTQFLMDHDOEP 218

363 DEPLHNGKQOQOESCEPELQVKEKQESPEDELHNGKQOQOESCEPELQVKEKQESPEDELHNGKQOQOESCEPEL 422

219 DKTGLYSKRAAKSDTSDDFMEGEGEDGSGDGGSEFLQRFSEYERHTES 278

423 ELHLG-----KQESHERPTEDEKQESLYEPELHLGKQOQOESCEPELQVKEKQESPEDELHNGKQOQOESCEPEL 472

279 LQNSKQOELIKYLELEKLSRMEDENNRL--RLESKRLGSDAVRELELRLRAEN 336

473 LMQLKQEKASRGQELDD--SHLEQEKELLDRLQDELVNKDE-----QLERKKHKLLEN 524

337 L-----QLTENEHLRQOERAPL 354

525 LQKQEKQIKQLVPSPD--RVQETQPI 548

Db

RESULT 4

A60746

chromogranin A precursor - rat

N:Contains: beta-granin; pancreastatin

C:Species: Rattus norvegicus (Norway rat)

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #ext\_change 21-Jul-2000

C:Accession: A60746; S00291; S02543

R:Parmer, R.J.; Koop, A.H.; Handa, M.T.; O'Connor, D.T.

Hypertension 14, 435-444, 1989

A:Title: Molecular cloning of chromogranin A from rat pheochromocytoma cells.

A:Reference number: A60746; MUID:90007662; PMID:2793216

A:Accession: A60746

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-462 <PAR>

A:Cross-references: GB:AF145445; NID:95163367; PIDN:AAA0652.1; PID:95163368

R:Iacangelo, A.; Okayama, H.; Elden, L.E.

FEBS Lett. 227, 115-121, 1988

A:Title: Primary structure of rat chromogranin A and distribution of its mRNA.

A:Reference number: S00291; MUID:88112232; PMID:2828116

A:Accession: S00291

A:Molecule type: mRNA

A:Residues: 1-107, 'QOQO', 108-462 <IAC>

A:Cross-references: EMBL:X06832; NID:955950; PIDN:CAA29988.1; PID:955951

R:Hutton, J.C.; Nielsen, E.; Kastern, W.

FEBS Lett. 236, 269-274, 1988

A:Title: The molecular cloning of the chromogranin A-like precursor of beta-granin and F

A:Reference number: S02543; MUID:88312980; PMID:3044025

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

A:Accession: S02543

Query Match 7.0%; Score 133; DB 1; Length 538;

Best Local Similarity 20.9%; Pred. No. 0.83;

Matches 89; Conservative 40; Mismatches 142; Indels 154; Gaps 19;

30 PERPGA-----EERPEEDSRMOSRAFPOLGRRPGE---GEGS-----LESOPPL 74

107 PPLPSPAPSLPLEPSTPRSSLYPAL-PSLGAKPKQVLSDSGLIDLTEDPPY 165

75 QTOACESSCLREGKEGNGDDSSAGGDPPEAEVPEPTAEELL---AOPCHDEASKL 130

166 RDRPPPSD-RDG-----NGEATPAGEAP-----DPSPMASRLRGRRPPVADSTTSDA 214

131 GAPAGGEEMGOQOLGKKRRRPSKKRRKRYVLTWEKKKPEKXSLASRR 190

215 PPLRAGN-----GOLTWPPSSDLYNWK-----NNPFSFSDPKLTALIE 257

191 AEMFAQGVAPVNTTQPLMDHDEEPDLKTG-----LYSKRAAKSDD----- 235

258 SVLI-----THQPIWDDCOQLGLTLTGEKQVLTLEARKAVRGGDGRPOLPNE 307

236 -----TSDDPFMEEGEE-----DGSDMGCGDSE 261

308 VDAAPFLERPDMDYTTQAGRNLVHYROLLAGLQAGRSPTNIAKVGITGCPNESPSA 367

262 FLORDSEPTERY-----HTESLONMSKOLLEKE 290

368 FLER-LKEAVRTTYPDPEDPGOETVWSMFTWQASPDIGRLGLEDLKNTLGLDLV- 425

291 YLELEKSLRMEDENNRLLESKRLGDDPARV-RELELDRLKAENLQLTENELHROQ 349

426 --EAEKIFPKRETPER-----EERLRRETEKEERRTEDGKEKERDORRRR 472

350 ERAPL 354

473 EMSKL 477

RESULT 6

A44224

DNA repair helicase ERCC6 - human

C:Species: Homo sapiens (man)

C:Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 21-Jul-2000

C:Accession: A44224

R:Truelstra, C.; van Gool, A.; de Wilt, J.; Vermeulen, W.; Bootsma, D.; Hoeijmakers, J

Cell 71, 939-953, 1992

A:Title: ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne

A:Reference number: A44224; MUID:93092214; PMID:1339317

A:Accession: A44224

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-1493 (CRO)

A:Cross-references: GB:I04791; NID:9182180; PIDN:AAA52397.1; PID:9182181

A:Note: sequence extracted from NCBI backbone (NCBIF:120238)

C:Keywords: DNA binding

Query Match 7.0%; Score 133; DB 2; Length 1493;

Best Local Similarity 23.7%; Pred. No. 2.3; 118; Indels 104; Gaps 17;

Matches 82; Conservative 42; Mismatches 118; Indels 104; Gaps 17;

35 GAERVP-----EEDSRMOSRAFPOLGRRPGEHGSLEOPPLQTOACESSCLREG 88

201 GAEVKIEDHSLSEDA-----EPGP-----SSLSMLMPQETAMEE--LIRKG 243

89 EKGNGDDSSAGGDPPEAEVPTP---EA-----ELLAPCHDEASK 129

244 QM-----TPFGTOIPQKQKKPRKIMLEASGFERYLADQAKLSFERKKQCNRAAK 297

130 LGAPA-----AGGEEMGOQOLGKKRRRPSKKRRKRYVLTWEKKKPEKXSL 184

298 AAPAVTPPAVONKKNPKARVLSKEER-----LKKHITKLQKRALQFQKVGILPKR 352

185 RA-----SRIIRAFMAKGVAPVNTTQPLMDHDEE---EPDL-----KYG 223

RESULT 5

FOVMIM

gag polypotein - Moloney murine leukemia virus

N:contains: core protein p15; core shell protein p30; inner coat protein p12; nucleoprotein

C:Species: Moloney murine leukemia virus

A:Note: host Mus sp. (mouse)

C:Date: 01-Sep-1981 #sequence\_revision 27-Nov-1985 #text\_change 24-Jul-1997

C:Accession: A03930

R:Shinnick, T.M.; Lerner, R.A.; Sutcliffe, J.G.

Nature 293, 543-548, 1981

A:Title: Nucleotide sequence of Moloney murine leukaemia virus.

A:Reference number: A93265; MUID:82035843; PMID:6169994

A:Accession: A03930

A:Molecule type: genomic RNA

A:Residues: 1-538 (SHI)

A:Experimental source: clone pMLV-1

C:Comment: This protein is synthesized as a gag-pol polypotein.

C:Genetics:

A:Gene: gag

C:Superfamily: mammalian retrovirus gag polypotein I

C:Keywords: core protein; inner coat protein; nucleoprotein; polypotein

F:2-131/Product: core protein p15 #status predicted <C15>

F:132-215/Product: inner coat protein p12 #status predicted <C12>

F:216-478/Product: core shell protein p30 #status predicted <C30>

F:479-534/Product: nucleoprotein p10 #status predicted <C10>

Db 353 RPWESMDRPEABECSDEGE -SEYFPTBEEBEEDEDEVEGACADJLSGDGTVELKPLPKGG 4111

Oy 224 LYKRAAKASDDTSDDDFMEEGEE -DGSDDMGSDGSEFLÖRPFSETHRYHTESJÖNN 2822

Db 412 KRÖKKYVÖE---IDDDFFPSSGEEAEAAVSVEGGGGGRKKRGRRDDGDEDEYVYKORLRW 468

Oy 283 SKÖELILEYLELEKLSRMEDENNNLRLIESKRLGSDRVARELE 328

Db 469 NK-----LRJÖDEKRLKLE-----DDESESDAEFD 494

RESULT 7  
JC7168  
lens epithelium-derived growth factor - human  
N:Alternate names: transcriptional co-activator p75 protein  
C:Species: Homo sapiens (man)  
C:Date: 04-Mar-2000 #sequence\_revision 04-Mar-2000 #text\_change 11-May-2000  
C:Accession: JC7168  
R:Singh, D.P.; Ohguro, N.; Kikuchi, T.; Sueno, T.; Reddy, V.N.; Yuge, K.; Chylack Jr., I.  
Biochem. Biophys. Res. Commun. 267, 373-381, 2000  
A:Title: Lens epithelium-derived growth factor: Effects on growth and survival of lens e  
A:Reference number: JC7168; MUID:20090645; PMID:10623637  
A:Accession: JC7168

A:Cross-references: GB:AF063020  
A:Experimental source: lens epithelial cells  
C:Comment: This protein, a regulatory factor, belongs to a family of growth factor, which on of this factor stimulates cell attachment, shortens the initial lag in growth, and e  
C:Genetics:  
A:Gene: ledgf  
C:Keywords: eye lens; glycoprotein; growth factor; growth regulation

Query Match	6.9%	Score 132.5;	DB 2;	Length 530;
Best Local Similarity	21.2%;	Pred. No. 0.87;		
Matches 77;	Conservative 56;	Mismatches 115;	Indels 115;	Gaps 17;

```

QY      7 SEYOHQOTNSCTGAAYVOELINERPPGAEEVRPEEDSRMOSARFQ-----L 55
Db      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      93 SSOQAATKQKOSMASSDVEVEEK-----ETSKEDTDHEKASNEVDYTKAVDITPK 143
Db      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      56 GGRPG--PEEGSGLESQPPLOTQACPESSCLREGEKGQNGDDSSAGDPPPPAEV--E 110
Db      144 AARGRRKRKAQEVETEELAGVTTATASVNLKVSPPKGR-----PAATEVWIPK 192
QY      111 PTPAEALLAQRCHHSESAKLGAPRAGGEEMGGQQRQLGKKKRRRRSKKRRHKPKYKL 170
Db      193 PRGPRKWKQPC--PSESDII-----TEEDSKSRKGGQEKQPKR----- 229
QY      171 TWEKKKKFDEKOSLRASRIRAFEMFAKQPVARYNTQTQLMDHDHOEPDLKTG---LYSK 227
Db      230 ---QPKKDEBQK-----BEDKPRKPPDKKEGKKEVESK 260
QY      228 RA-AAKSDDTSDDDFMEEGGEEDGSDGMGGDGESEFLORDFSEYER-----YHTESLQ 280
Db      261 RKNLAKTCGVSTSGSEEGDDQDEBEKKRKGG-----RNPQYARRNRRLKGQHEKRAAD 313
QY      281 NMSQOEILKEYLELEKCLSRMEDENNR---LRLESKRLGDDARVLELEL--DRLEAEN 336
Db      314 RKRKQE---EQMETEQ---QNKDEGKPRPEVKYKRETSMDSPHQRIHAIEIKNSLKDIN 367
QY      337 LQL 339
Db      368 LDV 370

```

RESULT 8  
E59436  
KIAA1314 protein [imported] - huamm  
C:Species: Homo sapiens (man)  
C:Date: 03-Jun-2002 #sequence\_revision 03-Jun-2002 #text\_change 19-Jul-2002  
C:Accession: E59436; F59436

R:Nagase, T.; Kikuno, R.; Ishikawa, K.I.; Hirose, M.; Ohara, O.  
DNA Res. 7, 65-73, 2000

A:Title: Prediction of the coding sequences of unidentified human genes. XVI. The com

A:Accession: E59436

A:Reference number: E59436

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-681 <NAG>

A:Cross-references: GB:BAA92552; PID:g7243009; PIDN:BAA92552.1

R:Ohara, O.; Nagase, T.; Kikuno, R.  
submitted to GenBank January 2000

A:Reference number: F59436

A:Accession: F59436

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-681 <NAG>

A:Cross-references: GB:BAA92552; PID:g7243009; PIDN:BAA92552.1

Query Match	6.9%	Score	131	DB	2	Length	681
Best Local Similarity	26.4%	Pred. No.	1.4				
Matches	84	Conservative	44	Mismatches	110	Indels	80
						Gaps	18

OY	66	SLEOPPLTOAQPESSC-----LREBKQONDDSSAG-dPPAEVPEPTEAE	116
Dd	14	SNEIHPAFARSNSEASVDASMEDFWRETETK---DSMGQGEPPEAEPVPEGE	70
OY	117	LILAOPCHDESKEKIGAPACGEPFEGGOOOLCKKKHRRRPSKKKHMKRYUKLTWEKK	176
Dd	71	LEAWMIDOVGLSTL---ISGDEEDGC--KALLSTLTRTQAAYOKR---YHTYTQTMRK	121
OY	177	KFEKOSLRASRIRAEFPAKQ-----PVARYNTQFLMDHDDEPDL-----KTGLYSK	227
Dd	122	K--DKQSITRDV---DIEFGVSESPRDTGSGNTHQL---GGTKEEERLPVYIKTSGMPD	173

```
QY 228 RAAKSODTSPDDMEGGEDGSDGMG-----GDGS-----EF 262
Db 174 DASLNSTTLLSDASODKSEFAPRSDSVALLETIPVLPHVSNSGSPBGPQYQNASIDDF 233
QY 263 LORD-----FSELYERYHTESLQ--NMSKOELLKEYLELJCKLSRMEDENRRL- 310
Db 234 LEKNIPEAELELSEFVSYSEKVTALRKKKLSEIKKE---DYVLTKFNVOCTRGTLT 289
QY 311 ESKRLGDD-ARVRELEL 327
Db 290 EAGDLSEAEDMKKIRHLSEL 307
```

```

RESULT 9
T00329
A: hypothetical protein KIAA0553 - human (fragment)
C: Species: Homo sapiens (man)
C: Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C: Accession: T00329
R: Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara,
DNA Res. 5, 31-39, 1998
A: Title: Prediction of the coding sequences of unidentified human genes. IX. The comp
A: Reference number: Z14086; MUID:98290545; PMID:9628581
A: Accession: T00329
A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: mRNA
A: Residues: 1-1095 <NAG>
A: Cross-references: EMBL:AB011125; NID:g3043629; PIDN:BAK25479.1; PID:g3043630
A: Experimental source: Brain
C: Genetics:
C: Note: KIAA0553

```

Query March 6.9% Score 131; DB 2: Length 1095;  
Best Local Similarity 19.7%; Pred. No. 2.2;  
Matches 99; Conservative 41; Mismatches 137; Indels 226; Gaps 21;

OY 13 POTSNTGGAAYVDLNPFGAEERYVP--EDSRMQSRAFPOLUGRGPeeEGE--- 65  
| : : : | | | | | : : : |  
Db 33 KPKSKCAKAAASQ-----GAERKVSESDPCKPTMTPE--SEPSSKAENAKALG 80



submitted to the EMBL Data Library, July 1996  
A:Description: Sla2 homologue of Yarrowia lipolytica.

A:Reference number: Z20763

A:Accession: T30177

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1054 <GCU>

A:Cross-references: EMBL:U65409; NID:93978133; PID:93978134; PIDN:AAC83184.1

A:Experimental source: strain W29

C:Genetics:

A:Gene: SLA2

A:Map position: III, adjacent to MATA

Query Match 6.8%; Score 129; DB 2; Length 1054;  
Best Local Similarity 23.4%; Pred. No. 2.8;

Matches 92; Conservative 45; Mismatches 124; Indels 132; Gaps 19;

31 ERPPGAEERVPEDSRMSRAFPOLGGRPEGE-GSLESOPPLQTOACESSCLREG 89

262 KNPNNLE---EDG-----PGLPQPKSVATRSVSPAPLPV-ATPTPSI----- 305

90 KGQNDSSAGDPPPAVEEPT---PEALLAQPCHESSKLGAPAAAGEEEMQ-- 143

306 -----PAENQPIEFMSQDALLAQOYDAQERLRQQAHEERIRQM 349

144 -QOROLGKKHRRPRSKKKRHKPYKLTWEKKKFE---KSLRASRTIEMFAKQ 198

350 MQOQOOFEMOQROQMEQORA-----OEOQIMADQMANHAGRMALERDLALRG 400

199 PVAAYNTQPLMDHDEERPLTGKLYSKRAAASDOTSDDFMEEGEGGSDMGD 258

401 ---YDDQMLLEEDYDRVKALELLELQLOQTAAQSAQKDLIES----- 442

259 GSEFLQDFSETRYTESLQNMKSQELIKEYLELEKCL-----SRM 301

443 ---LQOQITMWRKQYETLAKRYSSMR---EYLALKLKATQQAASAKALEKAKM 495

302 E-----DENNRILESKRL-GGDDARVLELEL-----DLRAEN 336

496 ERDRHKNIELADLIKERDRARYDLDRAGKGNKEDVERLERLMAQDLKDRSTGAD 555

337 LQLL-----TENELHQOERAPLSKRGD 359

556 LSLLSKHNRELSELENAL-KMKORA-LDERGD 586

RESULT 13

B45344

probable nuclear antigen - suid herpesvirus 1 (strain Kaplan)

C:Species: suid herpesvirus 1

C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 16-Jul-1999

C:Accession: B45344

R:Vicer, C.; Kozmik, Z.; Paces, V.; Schirm, S.; Schwytzer, M.

Virology 179, 365-377, 1990

A:Title: Pseudorabies virus immediate-early gene overlaps with an oppositely oriented or

A:Reference number: A45344; MUID:91021039; PMID:2171211

A:Accession: B45344

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-1733 <VIC>

A:Cross-references: GB:M4651; NID:9334070; PIDN:AAA47471.1; PID:9334072

C:Superfamily: pseudorabies virus 1 nuclear antigen

Query Match 6.7%; Score 128; DB 1; Length 1733;

Best Local Similarity 22.4%; Pred. No. 5.4;

Matches 74; Conservative 27; Mismatches 94; Indels 136; Gaps 16;

8 EYOHQOTSCT-----GAAYOE-----INPER-PPGAEER----- 39

104 DQHRPPPTTTTITKDPHODPLLPRTLQDEEDHLLRPTDRPSAKTNHNDPPG 163

40 VPEDDSWQSRAPFOLGGRGP-----EGESLESOPPLQ----- 75

164 GPPSTSHHHNQDPGGGPPPPRSTSSSSHQGPSTTRPPQRRPPRPPSPQKI 223

76 --TOACPESSCLR-----EGEKQNGDDSSAGS-----PEPPAEVEPTPEA 115

224 SETRAGSENTAQTILFSHSEKTLFHPMEGEGE-GDRGTAGGEDRDPPPPPPPPPP 282

116 ELLAQPCHESSKLGAPAAAGEEEMQOQOROLG---KKHRRPRSKKKRHKPYKLTW 172

283 PLPPPPPPPPPPQ---PPPAGCS---ARRRRGGGPPGRRGRRGRRR----- 326

173 EKKKPFDEKQSLRASRTIEMFAKQGVYAPYNTQPLMDHDEERPLTGKLYSKRAAK 232

327 -----AEGTEAA-----AADAEEEDGD----- 344

233 SDTSDDDFMEEGEGEGSD--GMGGDGE 261

345 GDEDEDDEDRAGBEGREDOGEGPRGAGGAGE 375

RESULT 14

T06310

hypothetical protein F11C18.80 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 22-Oct-1999

C:Accession: T06310

R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.;

ewes, H.W.; Mayer, K.F.X.; Scheller, C.

submitted to the Protein Sequence Database, April 1999

A:Reference number: Z15589

A:Accession: T06310

A:Molecule type: DNA

A:Residues: 1-852 <BEV>

A:Cross-references: EMBL:AL049607; GSPDB:GN00062; ATSP:F11C18.80

A:Experimental source: cultivar Columbia; BAC clone F11C18

C:Genetics:

A:Gene: ATSP:F11C18.80

A:Map position: 4

A:Introns: 34/3; 102/3; 205/3; 269/3; 282/3; 564/3; 602/3; 622/3; 639/3; 667/3

Query Match 6.7%; Score 127.5; DB 2; Length 852;

Best Local Similarity 23.2%; Pred. No. 2.8;

Matches 88; Conservative 51; Mismatches 145; Indels 95; Gaps 18;

1 MAEPFLSEYOHQOTSCTGAAYOEELNPERPPGAEERYPEEDSRMSRAFPOLGGRG 60

439 MAEE-LPEQSVPKKTANQKKKESSTEEVKPSASITATEE-VSEBNPTE---PQVTKKSG 492

61 PEGGSLSEQPP-PLTOACPESSCLREGEKGNGDDSSAGDPPPAVEVP-----TPEA 115

493 KVAASSSKTKRPTVPSSKSTSEKTVARQOSEKKVGVSDMAOESTPKPEKKRPGKKAIDE 552

116 ELLAQPCHESSKLGAPAAAGEEEMQO-----OQROLGKKHRRPR---SK 159

553 ESLHTSGDNEKPAVSSGKTLASKSKTKAKOTVESPSNKTNRKRSLOQGAAGSELVGR 612

160 KRRHW---KPYKLTWEE---KKF-----DEKQSLRASRTIEMFAKQGVAPYNTQ 207

613 IKVWPMDOAYVKGVESYDAAKKHIVYDDQOEI-----LVKKQKMSPLDESE 664

208 FLMD---DHQEPDLKT---GLYSKRA---AAKDDPTSDDD----- 240

665 LSQDEAADOTGQEDNSTVGSAGSKAAATPASKSSKTSQDKTASKSKDSKDSKRE 724

241 --FMEEGCEEDGSDGSDGSEFLQDFSETYERYTESLQNMKSQELIKEYLELEKCL 298

725 EASSEESSEEPKTYGKSGSSKSKNDIS-----SVKSGK-----SKAS 765

299 SRMEDENNRLESKRLOG 317

766 SKKKEEPSKATYSSKSKSG 784

RESULT 15





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 10:09:56 ; Search time 24 Seconds  
(without alignments)  
620.417 Million cell updates/sec

Title: US-09-972-758a-2  
Perfect score: 1910  
Sequence: 1 MAEPPLSEYOHOPOTSNGT.....LTENELHROOERAPLSKFGD 359

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt.40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	length	ID	Description
1	164.5	8.6	1224	1 DYNB.CHICK	P35458 gallus gall
2	139	7.3	756	1 BRS1.BOVIN	006002 bos taurus
3	137	7.2	536	1 GAG.MLVCH	P27460 cas-br-e mu
4	134.5	7.0	1181	1 NKX1.RAT	Q942m6 rattus norv
5	134	7.0	568	1 INVO.RAT	P48998 rattus norv
6	133	7.0	538	1 GAG.MLVMO	P03332 moloney mur
7	133	7.0	1493	1 ERRC6.HUMAN	003468 homo sapien
8	131	6.9	1898	1 TRHY.HUMAN	007283 homo sapien
9	130	6.8	466	1 CMGA.RAT	P10354 rattus norv
10	129.5	6.8	1816	1 AFG.HUMAN	P55196 homo sapien
11	128	6.7	1733	1 VNUA.PRYKA	P33485 pseudorabie
12	128	6.7	2476	1 ATRX.MOUSE	P30427 rattus norv
13	127.5	6.7	4687	1 PLEI.RAT	P30427 rattus norv
14	126.5	6.6	1407	1 TRHY.RABIT	P37709 oryctolagus
15	125.5	6.6	1549	1 TRHY.SHEEP	P27793 ovis aries
16	125	6.5	2004	1 MOZ.HUMAN	Q92794 homo sapien
17	124	6.5	449	1 CMGA.BOVIN	P05059 bos taurus
18	123.5	6.5	677	1 SGL.MOUSE	P16014 mus musculu
19	122.5	6.4	4473	1 PLEI.CRIGR	Q9155 cricetus
20	122	6.4	723	1 SSRP.DROME	Q03344 drosophila
21	121	6.3	538	1 GAG.MSVMO	P03334 moloney mur
22	121	6.3	726	1 BRD3.HUMAN	Q15059 homo sapien
23	121	6.3	1089	1 Y553.HUMAN	Q94X13 homo sapien
24	121	6.3	1280	1 DYNB.RAT	P28023 rattus norv
25	120	6.3	536	1 GAG.MLVDE	P29168 rattus norv
26	120	6.3	699	1 NP14.HUMAN	Q14978 murine leuk
27	120	6.3	1781	1 ARAQ.HUMAN	Q02952 homo sapien
28	119.5	6.3	556	1 HIR3.HUMAN	Q9Bw71 homo sapien
29	118.5	6.2	472	1 TWIE.CAEEL	Q23525 caenorhabdi
30	118.5	6.2	764	1 UBFI.RAT	P25977 rattus norv
31	117.5	6.2	372	1 TOLA.HAEIN	P44678 haemophilus
32	117.5	6.2	513	1 DMP1.HUMAN	Q13316 homo sapien
33	117.5	6.2	2142	1 BAT2.HUMAN	P48634 homo sapien

34	117	6.1	568	1 RGS3.MOUSE	Q9dC04 mus musculu
35	117	6.1	845	1 NFM.RAT	P12839 rattus norv
36	117	6.1	911	1 CAFI.MOUSE	Q9qW10 mus musculu
37	116.5	6.1	503	1 DMP1.MOUSE	Q55188 mus musculu
38	116.5	6.1	1281	1 DYNB.MOUSE	008788 mus musculu
39	116	6.1	927	1 YOL3.CAEEL	Q02328 caenorhabdi
40	116	6.1	1729	1 TABP.HUMAN	Q9C0C2 homo sapien
41	116	6.1	1972	1 MYHB.RABIT	P35748 oryctolagus
42	115.5	6.0	793	1 CALD.HUMAN	Q05682 homo sapien
43	115.5	6.0	2472	1 NCR2.MOUSE	Q9W42 mus musculu
44	115	6.0	1739	1 CHD2.HUMAN	Q14647 homo sapien
45	115	6.0	1745	1 ZOL.MOUSE	P39447 mus musculu

## ALIGNMENTS

RESULT 1  
ID DYNB.CHICK STANDARD: PRT: 1224 AA.  
AC P35458:  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Dynactin 1 (150 kDa dynein-associated polypeptide) (DP-150) (DAP-150)  
DE (p150-glned).  
GN DCTNL.  
OS Gallus gallus (chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA CleveLand D.W.;  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 172-1224 FROM N.A.  
RC TISSUE=Embryonic brain;  
RX MEDLINE=92098576; PubMed=1836789;  
RA Gill S.R., Schroer T.A., Szilak I., Steuer E.R., Sheetz M.P.,  
RT "Dynactin, a conserved, ubiquitously expressed component of an  
RT activator of vesicle motility mediated by cytoplasmic dynein.";  
RL J. Cell Biol. 115:1639-1650(1991).  
CC -!- FUNCTION: DYNACTIN IS A MAJOR COMPONENT OF ACTIVATOR I. A 20S  
CC POLYPEPTIDE COMPLEX THAT STIMULATES DYNEIN-MEDIATED VESICLE  
CC TRANSPORT.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS ARE PRODUCED BY  
CC -!- ALTERNATIVE SPLICING.  
CC -!- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED.  
CC -!- SIMILARITY: CONTAINS 1 CAP-GLY DOMAIN.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
DR EMBL: X62773; CA44617.2; -.  
DR PIR: A41642; A41642.  
DR InterPro: IPR000938; CAP-Gly.  
DR Pfam: PF01302; CAP-Gly; 1.  
DR PROSITE: PS00845; CAP-Gly\_1; 1.  
DR PROSITE: PS0245; CAP-Gly\_2; 1.  
KW Motor protein; Microtubules; Dynein; coiled coil; Cytoskeleton;  
KW Alternative splicing.  
FT DOMAIN 49 91 CAP-GLY.  
FT DOMAIN 205 540 COILED COIL (POTENTIAL).

FT DOMAIN 936 1042 COILED COIL (POTENTIAL).  
 FT DOMAIN 1081 1117 COILED COIL (POTENTIAL).  
 SQ SEQUENCE 1224 AA; 135562 MW; 03B7FEE68E7C01D7 CRC64;

Query Match  
 Best Local Similarity 22.6%; Pred. No. 0.027; DB 1; Length 1224;  
 Matches 83; Conservative 61; Mismatches 152; Indels 71; Gaps 14;

29 NPERPPGAEEYRPEDSRWQSRAPOLGGRPGEGESLSQPPLOTQACPESSCLREG 88  
 106 SPETPESALAKVPRKHSRAAKGSKLRGAKP-----KTTARPKPTRTPTSA PSS-----G 157  
 QY 89 EKGQNDSSAGGDP--PPPAVEPTP-EAELLAPCHDSKSGKLGAPAGGEEBEGQOO 145  
 DB 158 TAGPSGSASASGSESSSEPTPAOTPLVAPVPSPLSPVAPVPSPTKEENLRSGV 217  
 QY 146 ROLGK-----KHRRRPSKKRHWKPYKLTWEE-----KKKPEKOSLASRR-LEMFAK 196  
 DB 218 RDLERKLETLKTRNEDKAKLLEKRYKIQLEOVEMKSKMOQOADIORLKEAKKRAK 277  
 QY 197 GQVAPYNTQFLMDHDQ-EEPDLTGLYSKRAAKSDDTSDDDFMEEGCE----- 247  
 DB 278 DALEAKERMEEMATADALEMATLDKEMAEERASLQEEV--DSLKEKVELTMDLEI 334  
 QY 248 -----EDGSDGMDGSEPLORDPSETYERYHTESLONMSKOEIik-----EYL 292  
 DB 335 LKHEIEEKSGDCA-----ASSYOVKOLEQONARKLEALVRMDLSASEKOEHV 382  
 QY 293 ELEKCLSRMEDENNRLRLESKRIGGDDARVRELLELRLR-----AENIOLLLEN 343  
 DB 383 KIQKMEKKNTELESRLROGREKL---QEEVKOAEKTVBELKQVADALGAEEMVELTER 439  
 QY 344 ELHROOE 350  
 DB 440 NLDLEEK 446

RESULT 2  
 BFS1\_BOVIN STANDARD: PRT; 756 AA.  
 AC Q06002;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Filensin (Beaded filament structural protein 1).  
 GN BFS1.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Lens;  
 RX MEDLINE=93260017; PubMed=8491777;  
 RA Gounari F., Merdes A., Quinlan R., Hess J.F., Fitzgerald P.G.,  
 RA Ouzounis C.A., Georgatos S.D.;  
 RT "Bovine filensin possesses primary and secondary structure similarity  
 RT to intermediate filament proteins.";  
 RL J. Cell Biol. 121:847-853(1993).  
 RN [2]  
 RP REVISIONS, AND SEQUENCE FROM N.A.  
 RC TISSUE=Lens;  
 RA Hess J.F.;  
 RA Submitted (DEC-1997) to the EMBL/Genbank/DBJ databases.  
 CC -I SUBUNIT: ASSOCIATES WITH BFS2.  
 CC -I SUBCELLULAR LOCATION: MEMBRANE- AND CYTOSKELETON-ASSOCIATED;  
 CC -I TISSUE SPECIFICITY: LENS.  
 CC -I SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X72388; GAA51081.1; -;  
 DR EMBL; X72388; GAA51081.1; -;  
 DR InterPro: IPR001664; IF.  
 DR PROSITE: PS00226; IF; FALSE\_NEG..  
 KW Intermediate filament; Repeat; Membrane; Coiled coil; Cytoskeleton;  
 KW Eye lens protein; Phosphorylation;  
 FT DOMAIN 1 38 HEAD.  
 FT DOMAIN 39 318 ROD.  
 FT DOMAIN 319 755 TAIL.  
 FT DOMAIN 39 73 COIL 1A.  
 FT DOMAIN 74 82 LINKER 1.  
 FT DOMAIN 83 182 COIL 1B.  
 FT DOMAIN 183 199 LINKER 12.  
 FT DOMAIN 200 318 COIL 2.  
 FT DOMAIN 531 621 7 x 14 AA TANDEM REPEATS.  
 FT DOMAIN 531 544 2 (INCOMPLETE).  
 FT REPEAT 545 551 3.  
 FT REPEAT 552 565 2.  
 FT REPEAT 566 579 4.  
 FT REPEAT 580 593 5.  
 FT REPEAT 594 607 6.  
 FT REPEAT 608 621 7.  
 FT MOD\_RES 5 5 PHOSPHORYLATION (BY PKA) (POTENTIAL).  
 SQ SEQUENCE 756 AA; 83001 MW; F86A18208A8E109 CRC64.

Query Match  
 Best Local Similarity 22.6%; Pred. No. 0.44;  
 Matches 88; Conservative 32; Mismatches 117; Indels 152; Gaps 16;

26 FELNPERPPGAEEYRPEDSRWQSRAPOLGGRPGEGESLSQPP-PIQTOA----- 78  
 DB 415 EEASPTTQEGAPEDVP--DGCKISKAEFLKMKIKKVKGRKEPPADLTTKGRYVWS 472  
 QY 79 -----CPSSCLREGKEKGNGDDSDSAGDPFPPPAVEPTP-----EAELLAQ 120  
 DB 473 GDGSFVDPGFCVFSVPKAGVAVSGDSDS-----VPPDSGVPEPSPQDEPLEEGQGPQ 527  
 QY 121 PCHSSEASKLGAPAAAGP----- 587  
 DB 528 EKEGGLKEEGPPEEGKGPPEGGKDDSVKKEGPGPEGGKGVKKEGPGKGVKKEG 587  
 QY 143 QOO-ROLGKKKKHRRRPSKKRHWKPYKLTWEEKKKPEDEKOSLRASRLRABMFAKGPVA 201  
 DB 588 PREGKGDGVKKEGPEPEGGKGLK-----EEGPIQDEKD-----GOSPT 627  
 QY 202 PYNTQFLMDHDQEEPDLTGLYSKRAAKSDPTSDDDFMEG-----GEED- 249  
 DB 628 PHPA-----DKGDEKNAKELKLGLOGKO-----DDQKEGARGCPVAVAPGPGPS 672  
 QY 250 -----GSDGMDGSDSEFLQ-----DSEYTERHTESLONMSKOEI 287  
 DB 673 TPSSQGPQVILGSGEGHARSGRRLARSPPKLAEKVEKVESEIEKSTESIQYEETAV 732  
 QY 288 IKLEYLEKCLSRMEDENNRLRLESKRIG 316  
 DB 733 IVERFM-IEK-----TKANKKILG 749

RESULT 3  
 GAG\_MLVCB STANDARD: PRT; 536 AA.  
 ID GAG\_MLVCB  
 AC P27460;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE GAG polyprotein [contains: Core protein p15; Inner coat protein p12;  
 DE Core shell protein p30; Nucleoprotein p10].  
 GN GAG.  
 OS Cas-Bt-E murine leukemia virus.

OC Viruses: Retroviral viruses: Retroviridae: Gammaretrovirus.  
OX NCBI\_Taxid=111792;  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91227170; PubMed=1840655;  
RA Berryman S.M., McAtee F.J., Portis J.L.;  
RT "Complete nucleotide sequence of the neurotropic murine retrovirus  
CAS-BR-E";  
RL Nucleic Acids Res. 19:1707-1707(1991).  
CC -1- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES: IN VIVO YIELD MATURE PROTEINS.  
CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-POL  
CC POLYPROTEIN.  
CC -1- SIMILARITY: CONTAINS 1 CCHC-TYPE 2INC FINGER.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X57540; CAA40759.1; -  
DR InterPro: IPR000840; Gag\_MA.  
DR InterPro: IPR002079; Gag\_P12.  
DR InterPro: IPR003036; Gag\_P30.  
DR InterPro: IPR001878; Znf\_CCHC.  
DR Pfam: PF00098; zf-CCHC; 1.  
DR Pfam: PF01140; Gag\_MA; 1.  
DR Pfam: PF01141; Gag\_P12; 1.  
DR Pfam: PF02093; Gag\_P30; 1.  
DR SMART: SM00343; Znf\_C2HC; 1.  
DR PROSITE: PSF0158; ZF-CCHC; 1.  
DR Core protein: Coat protein; Nucleoprotein; Polypotein; Myristate;  
KM Zinc-finger.  
KW CHAIN 2 129 CORE PROTEIN P15.  
FT CHAIN 130 214 INNER COAT PROTEIN P12.  
FT CHAIN 215 477 CORE SHELL PROTEIN P30.  
FT CHAIN 478 536 NUCLEOPROTEIN P10.  
FT ZN\_FING 500 517 CCHC-TYPE.  
FT LIPID 2 MYRISTATE (BY SIMILARITY).  
SQ SEQUENCE 536 AA: 60740 MW: 3E0F863A393960AE CRC64;  
  
Query Match 7.2%; Score 137; DB 1; Length 536;  
Best Local Similarity 21.3%; Pred. No. 0.39;  
Matches 88; Conservative 54; Mismatches 141; Indels 130; Gaps 19;  
  
OY 30 PERPPAEERVPEDSRMOSRA-----FPQUGRGPPE---GEGS-----LESOPPPQ 75  
DB 105 PLPLPPAPSLPLPPPLSTPRSSLYPALTPSLGAKPKQVLPDPSGGPLIDLLEDPDPYR 164  
OY 76 TOACPPSSCLRECKQNGDSSAGCDPPPAVEPTPEALL-----AQPCHESEASKG 131  
DB 165 DPGPPS-----DRRDGGEAPAGGEAP-----DSPMASRLRGRELVPADSTISQAF 213  
OY 132 APAAGGEEMGOOQROLGKKHRRPSKKRHMK---PYV-----KL----- 170  
DB 214 PLRSGN-----GOLQYWPSSDLYNMKNPNFSEDPCKLTALIESVLTTHQ 263  
OY 171 TWEEKKKF-----DEKSLASRIRAEKMGQVAVYNTQFLMDHDOEPP----- 219  
DB 264 TWDDCOQLLTGTLTGEEKORV---LLEAKKAVGEGDGRPTQLPNEINDAFPLERPMDYN 320  
OY 220 -----LKTGLYSKRAAKSDTSDDDMEGCEGDESDMGDGSFFLOR 265  
DB 321 TQGRNHLVLYROLLAGLON---AGRSPTNL-----AKKGTIGQPNESPFAFLER 369  
OY 266 DPESEYERYHTESLQ-----NMSKQELIKYLELEKCLSRMEDENNRLLSESKRLGG-- 317  
DB 370 -LKEAVRRYPTDPDEPQGETVMSMFIWQASADIDIKLERLED-----LMSKTLGDLV 422  
OY 318 -----DDARVRELELELDRLRAENQLLTENFLHQGERAPL 354

DB 423 REAEKIFNKRRETPEREERIKETEKEERRAEDEQEKERDRRRHREMSKL 475  
RESULT 4  
ID NCX1\_RAT STANDARD; PRT; 1181 AA.  
AC Q9Q2M6; Q62932;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Sodium/potassium/calcium exchanger 1 precursor (Na(+)/K(+)/Ca(2+)-  
DE exchange protein 1) (Retinal rod Na-Ca+K exchanger).  
GN SLC24A1 OR NCX1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_Taxid=10116;  
RN [1]  
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
RC STRAIN=Sprague-Dawley; TISSUE=Eye;  
RX MEDLINE=20217335; PubMed=10751314;  
RA Poon S., Leach S., Li X.-F., Tucker J.E., Schneekamp P.P.M.,  
RA Lytton J.;  
RT "Alternatively spliced isoforms of the rat eye  
RT sodium/calcium-potassium exchanger NCX1.";  
RL Am. J. Physiol. 278:C651-C660(2000).  
RN [2]  
RP SEQUENCE OF 1067-1155 FROM N.A.  
RA White K.E., Gesek F.A., Friedman P.A.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: Critical component of the visual transduction cascade,  
CC controlling the calcium concentration of outer segments during  
CC light and darkness. Light causes a rapid lowering of cytosolic  
CC free calcium in the outer segment of both retinal rod and cone  
CC photoreceptors and the light-induced lowering of calcium is caused  
CC by extrusion via this protein which plays a key role in the  
CC process of light adaptation. Transports one Ca(2+) and one K(+) in  
CC exchange for four Na(+).  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1 (shown here) 2, 3 and 4; are  
CC produced by alternative splicing.  
CC -1- TISSUE SPECIFICITY: Highly expressed in the eye.  
CC -1- SIMILARITY: BELONGS TO THE SLC24A FAMILY OF TRANSPORTERS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AF176688; AAD53121.1; -  
DR EMBL: U49235; AAB37753.1; -  
DR InterPro: IPR004817; K\_NaCaexchang.  
DR InterPro: IPR004481; K\_NaCaexchang.  
DR InterPro: IPR004837; NaCa\_Exmemb.  
DR Pfam: PF01699; Na\_Ca\_Ex; 3.  
DR TIGRfams: TIGR00367; K\_NaCaexchang-rel; 1.  
DR TIGRfams: TIGR00927; ZAI904; 1.  
KW Vision; Transport; Antiport; Symport; Calcium transport;  
KW Transmembrane; Glycoprotein; Phosphorylation; Signal; Repeat;  
KW Alternative splicing.  
FT SIGNAL 1 38  
FT CHAIN 39 1181  
FT DOMAIN 39 419 SODIUM/POTASSIUM/CALCIUM EXCHANGER 1.  
FT TRANSMEM 420 440 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 441 464 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 465 485 POTENTIAL.  
FT DOMAIN 486 491 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 492 512 POTENTIAL.  
FT DOMAIN 513 519 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 520 544 POTENTIAL.

FT	DOMAIN	545	552	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	553	569	POTENTIAL.
FT	DOMAIN	570	989	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	990	1010	POTENTIAL.
FT	DOMAIN	1011	1017	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1018	1038	POTENTIAL.
FT	DOMAIN	1039	1053	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	1054	1074	POTENTIAL.
FT	DOMAIN	1075	1092	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1093	1113	POTENTIAL.
FT	DOMAIN	1114	1121	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	1122	1142	POTENTIAL.
FT	DOMAIN	1143	1150	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1151	1171	POTENTIAL.
FT	DOMAIN	1172	1181	CYTOPLASMIC (POTENTIAL).
FT	REPEAT	461	501	ALPHA-1.
FT	REPEAT	1061	1092	ALPHA-2.
FT	DOMAIN	730	905	14 X APPROXIMATE TANDEM REPEATS.
FT	REPEAT	730	741	1.
FT	REPEAT	742	754	2.
FT	REPEAT	755	766	3.
FT	REPEAT	767	778	4.
FT	REPEAT	779	791	5.
FT	REPEAT	792	804	6.
FT	REPEAT	805	817	7.
FT	REPEAT	818	830	8.
FT	REPEAT	831	843	9.
FT	REPEAT	844	856	10.
FT	REPEAT	857	869	11.
FT	REPEAT	870	881	12.
FT	REPEAT	882	893	13.
FT	REPEAT	894	905	14.
FT	DOMAIN	952	974	POLY-GLUT.
FT	MOD_RES	625	625	PHOSPHORYLATION (POTENTIAL).
FT	CARBOHYD	271	271	N-LINKED (GLUCNAIC. . . ) (POTENTIAL).
FT	VARSPLC	598	710	MISSING (IN ISOFORM 2).
FT	VARSPLC	616	710	MISSING (IN ISOFORM 3).
FT	VARSPLC	652	679	MISSING (IN ISOFORM 4).
SO	SEQUENCE	1181 AA:	129980 MW:	B063C1C1193696AE CRC64;

Query Match	7.0%;	Score 134.5;	DB 1;	Length 1181;
Best Local Similarity	22.5%;	Pred. No. 1.2;		
Matches	67;	Conservative	44;	Mismatches 160; Indels 27; Gaps 7;
QY	19	TGAANOVELMPERPSCAEERYVEDDSW----	QSRAFPOLGCRPGEGCSLESOPRP	73
Dp	692	TPAPAPREKSGOEEDPPCGOEDVDVAENHGDGMTGEGRETEAEGKKDEGETAEARKBDG	751	
QY	74	LQTAACPSSCLREGEKQONGDDSSAGSDPPRAPEVETPTPEALLAOC--	HDSEASKIG--	131
Dp	752	QOEETETKGRKQKOEFTPESEKDDOEGTTEAYGKADHGEETAEAGKEVEYHGETEAECT	811	
QY	132	-----APAAAGEEEMWGOOQOLKKKKHRRRRSSKKKHKMKPYKKLTWEKKKKFEDEKSL	184	
Dp	812	EDEOEGETEAGKEVEYEGGETEAGKEVEYHEVETEAEKRTKNHGEGETEAGKEADHGEGET	871	
QY	185	RAS---RIRAEPMFAGQVAPARYNTPTFLMDND-----	QEPRLKTLTKYKRAAKADD	235
Dp	872	EAEENVEHQGTTEAGKEVNEGETEAGCKDDHDEGOSQTQADDEKVD--	BGEAANAED	929
QY	236	TSDDDFMEEGEEDGSDMGDSEFPQRFSETEYKRYHNSLQMSKQDILKEYLE	293	
Dp	930	QCETAGCKGADGGGGSD--	CGDSPEDEDEDEEEEEEDEEESESEFPUSLEMP	985
RESULT 5				
INVO_RAT				
ID_INVO_RAT	STANDARD:	PRT:	568 AA.	
AC	P48998:			
DT	01-FEB-1996 (Rel. 33, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Involucrin.			

OS	IVL.	
OS	Rattus norvegicus (Rat).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OX	NCBI_TaxID=101116;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-Sprague-Dawley;	
RX	MEDLINE=9410476; Pubmed=8277848.	
RA	Djian P., Phillips M., Easley K., Huang E., Simon M., Rice R.H.,	
RA	Green H.;	
RT	"The involucrin genes of the mouse and the rat: study of their shared	
RT	repeats.";	
RL	Mol. Biol. Evol. 10:1136-1149(1993).	
CC	-1- FUNCTION: Part of the insoluble cornified cell envelope (CE) of	
CC	stratified squamous epithelia.	
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic. Constituent of the scaffolding	
CC	of the cornified envelope.	
CC	-1- TISSUE SPECIFICITY: Keratinocytes of epidermis and other	
CC	stratified squamous epithelia.	
CC	-1- PTM: Substrate of transglutaminase. Specific glutamines or lysines	
CC	are cross-linked to keratins, desmoplakin and to inter involucrin	
CC	molecules.	
CC	-1- SIMILARITY: BELONGS TO THE INVOLUCRIN FAMILY.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	-----	
DR	EMBL; L28818; AAA41445.1; -	
DR	InterPro; IPR002360; Involucrin.	
DR	PROSITE; PS00795; INVOLUCRIN; 1.	
KW	Keratization; Repeat.	
QO	SEQUENCE 568 AA; 67021 MW; E2B07740FE779345 CRC64;	

```

Query Match      7.0%; Score 134; DB 1; Length 568;
Best Local Similarity 20.9%; Pred. No. 0.61;
Matches 93; Conservative 66; Mismatches 169; Indels 118; Gaps 18;

QY      3 EPELSEYONOROTSNETGAAYOE-----LNPERPRAEERYEED 44
      1 : : : : : : : : : : : : : : : : : : : : : :
Db      127 EQLILYDHOHROFSESOGSLSGQDQDYLARQELMGNQOKELQERLPGQQQKTRBEQ 186
      1 : : : : : : : : : : : : : : : : : : : : : :

QY      45 -----SRMOSRATPOLLSGRGRPE-----GESLSERPRLTQAS-----PESSCLR 86
      1 : : : : : : : : : : : : : : : : : : : : : :
Db      187 ELLILGEKQOQLILVERHNOERQEBELNNGOKQOQROEYOLATOVONOKONERE--LCRL 245
      1 : : : : : : : : : : : : : : : : : : : : : :

QY      87 EGEKONGDDSSAGDPRPRAVEYER-----PREAEILAPRCHNSEASKLGAR 133
      1 : : : : : : : : : : : : : : : : : : : : : :
Db      246 KQOQOESHENELNLGKQOQOESHNERELNLGKQOQOESHNERELNLGKQOQOESEREL--- 302
      1 : : : : : : : : : : : : : : : : : : : : : :

QY      134 AAGSEEWGQOQOQO-----LGKKKNRR-----RPSKKKNRRKYULLEWKKKF- 178
      1 : : : : : : : : : : : : : : : : : : : : : :
Db      303 NLGEOQOHOEOQOQOESERELNLGKQOQOQOESLQKQOQYRNDYLDLRKQOKL 362
      1 : : : : : : : : : : : : : : : : : : : : : :

QY      179 -----DEKQSLKRSRLRAEM-----PAKQOVARUPTQRLMDNDHOER 218
      1 : : : : : : : : : : : : : : : : : : : : : :
Db      363 DRELNLGKQOQOQOESELQVERKKNQESRERELNLGKQOELNPRMTDEQEKQSLNPR 422
      1 : : : : : : : : : : : : : : : : : : : : : :

QY      219 DLKTGLYSKKAANKSDDTDDRMEEGSEGGSDGSGGSGSEFLQDRDSETERYNITES 278
      1 : : : : : : : : : : : : : : : : : : : : : :
Db      423 EHLNLG-----KQOESHPRDTEQOKQSLYERELNLGKQOQOQOEYEGYOR--SKS 472
      1 : : : : : : : : : : : : : : : : : : : : : :

QY      279 LQMSKQOELLKEYLEKSLSMEDENLRL--RLSKSLRGSDQARVYELDELDRPRAEN 336
      1 : : : : : : : : : : : : : : : : : : : : : :
Db      473 LNLGLKQEKASRQOELDD--SHLEQKELLDDQLDELVLNKDE-----OLEKKKKKL 524
      1 : : : : : : : : : : : : : : : : : : : : : :

QY      337 L-----QLTLENELNROOERAPL 354
      1 : : : : : : : : : : : : : : : : : : : : : :

```

Db 525 LTOKEKQIKQLYPSTD--RVQETQPI 548

RESULT 6

GAG\_MLVMO STANDARD; PRT; 538 AA.

ID GAG\_MLVMO STANDARD; PRT; 538 AA.

AC P03332;

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE GAG polyprotein [contins: Core protein p15; Inner coat protein p12; Core shell protein p30; Nucleoprotein p10].

GN GAG.

OS Moloney murine leukemia virus.

OC Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.

OX NCBI\_TaxID=11801;

RN [1]

RP SEQUENCE FROM N.A. (CLONE PHLV-1).

RX MEDLINE=82035843; PubMed=6169994;

RA Shindick T.M., Lerner R.A., Sutcliffe J.G.;

RT "Nucleotide sequence of Moloney murine leukaemia virus.";

RL Nature 293:543-548(1981).

RN [2]

RP SEQUENCE OF 2-31, AND MYRISTOYLATION.

RX MEDLINE=83169654; PubMed=6340098;

RA Henderson L.E., Kruttsch H.C., Oroszlan S.;

RT "Myristyl amino-terminal acylation of murine retrovirus proteins: an unusual post-translational proteins modification.";

RL Proc. Natl. Acad. Sci. U.S.A. 80:339-343(1983).

RN [3]

RP SEQUENCE OF 479-529.

RX MEDLINE=81264245; PubMed=6267042;

RA Henderson L.E., Copeland T.D., Sowder R.C., Smythers G.W., Oroszlan S.;

RT "Primary structure of the low molecular weight nucleic acid-binding proteins of murine leukemia viruses.";

RL J. Biol. Chem. 256:8400-8406(1981).

CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-POLYPROTEIN.

CC -1- SIMILARITY: CONTAINS 1 CCHC-TYPE ZINC FINGER.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: J02255; AAB59942.1; -

DR PIR: A03930; FOMVIM.

DR InterPro: IPR000840; Gag\_MA.

DR InterPro: IPR002079; Gag\_p12.

DR InterPro: IPR003036; Gag\_p30.

DR InterPro: IPR001878; ZnF\_CCHC.

DR Pfam: PF00098; zf-CCHC; 1.

DR Pfam: PF01140; Gag\_MA; 1.

DR Pfam: PF01141; Gag\_p12; 1.

DR Pfam: PF02093; Gag\_p30; 1.

DR SMART: SM00343; ZnF\_C2HC; 1.

DR PROSITE: PS50158; ZF\_CCHC; 1.

KW Core protein; Coat protein; Nucleoprotein; Polyprotein; Myristate; Zinc-finger.

KW Zinc-finger.

FT CHAIN 2 131 CORE PROTEIN P15.

FT CHAIN 132 215 INNER COAT PROTEIN P12.

FT CHAIN 216 478 CORE SHELL PROTEIN P30.

FT CHAIN 479 534 NUCLEOPROTEIN P10.

FT ZN\_FING 502 519 CCHC-TYPE.

FT LIPID 2 2 MYRISTATE.

SO SEQUENCE 538 AA; 60858 MW; 8A7652439B464495 CRC64;

Query Match 7.0%; Score 133; DB 1; Length 538;

Best Local Similarity 20.9%; Pred. No. 0.66;

Matches 89; Conservative 40; Mismatches 142; Indels 154; Gaps 19;

QY 30 PERPGA-----EERVPEDSRMQSRAPOLGGRGPE---GECS-----LESDPPPL 74

Db 107 PLPLPSAPSLPLEPPSTPPRSSLYPALT-PSLGAKPKQVLSDSGCLIDLTEDPPPY 165

QY 75 QTOQAPESSCLREGEGQNGDDSSAGDPPPAVEVEPTPEALL-----AQPCHDSEASKL 130

Db 166 RDRPPPSD--RCG-----NCGEATPAGCAP-----DPSPMASRLRCRRPPVADSTTSGA 214

QY 131 GAPAGGEEBEGQOQOLGKKHRRPSSKKRRHWKPYKLTWEKKKPDCKOSLRASRLR 190

Db 215 FPLRAGGN-----CQIQYWPSSSDLYNMK-----NNNSFSEDPKLTALIE 257

QY 191 AEMFAKGQVPAPVNTTQPLMDHDHGEPLKTG-----LYSKRAAKSND 235

Db 258 SVLI-----THQPTWDDCQQLLGTLLTGEKQRYVLEARKKAVRGDGGKPPQLPNE 307

QY 236 -----TSDDPFMEEGEE-----DGSDMGGDGSE 261

Db 308 VDAPFLERPDMDYTTQACGNHLVHYRQILLACLONAAGSPNLAKEVKGITGCPNESPFA 367

QY 262 FLDRPFSEYERY-----HRESIQNMKSQELIKE 290

Db 368 FLER-LKEAVRRYTPYDPEDPCQETNVMSFTWQAPDIGRKLERLEDLKNKTLGDLVR- 425

QY 291 YLELEKCLSMEDENNRRLRESKRLGGDARV-RELELEDRLRAENLQLTENELHROO 349

Db 426 --EAERIFNKRTEPEER-----EERIRRETEEKERRRTEDEQEKERDRRRHR 472

QY 350 ERAPL 354

Db 473 EMSKL 477

RESULT 7

ERC6\_HUMAN STANDARD; PRT; 1493 AA.

ID ERC6\_HUMAN STANDARD; PRT; 1493 AA.

AC 003468;

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Excision repair protein ERCC-6 (Cockayne syndrome protein CSB).

GN ERCC6 OR CSB.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OC NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=93092214; PubMed=1339317;

RA Troelstra C., van Gool A., de Wit J., Vermeulen W., Bootsma D., Hoeijmakers J.H.J.;

RT "ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes.";

RL Cell 71:939-953(1992).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=93181229; PubMed=8382798;

RA Troelstra C., Hesen V., Bootsma D., Hoeijmakers J.H.J.;

RT "Structure and expression of the excision repair gene ERCC6, involved in the human disorder Cockayne's syndrome group B.";

RL Nucleic Acids Res. 21:419-426(1993).

RN [3]

RP REVIEW ON VARIANTS CSB.

RX MEDLINE=99374920; PubMed=10447254;

RA Cleaver J.E., Thompson L.H., Richardson A.S., States J.C.;

RT "A summary of mutations in the UV-sensitive disorders: Xeroderma pigmentosum, Cockayne syndrome, and trichothiodystrophy.";

RL Hum. Mutat. 14:9-22(1995).

RN [4]

RP VARIANTS CSB, AND VARIANTS.

RX MEDLINE=98107940; PubMed=9443879;

RA Mallery D.L., Tanganelli B., Colella S., Steingrimsdottir H.,  
RA van Gool A.J., Troelstra C., Stefanini M., Lehmann A.R.;  
RT "Molecular analysis of mutations in the CSB (ERCC6) gene in patients  
RT with Cockayne syndrome.";  
CC Am. J. Hum. Genet. 62:77-85(1998).  
CC -1- FUNCTION: IS INVOLVED IN THE PREFERENTIAL REPAIR OF ACTIVE GENES.  
CC PRESUMED DNA OR RNA UNWINDING FUNCTION. CORRECTS THE UV SURVIVAL  
CC AND RNA SYNTHESIS AFTER UV EXPOSURE OF COCKAYNE'S SYNDROME  
CC COMPLEMENTATION GROUP B.  
CC -1- SUBUNIT: INTERACTS WITH THE CSB PROTEIN AND A SUBUNIT OF RNA  
CC POLYMERASE II TFIIH.  
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -1- DISEASE: Defects in ERCC6 are the cause of Cockayne's syndrome  
CC type B (CSB). CSB is a disease which is characterized by dwarfism,  
CC precociously senile appearance, pigmentary retinal degeneration,  
CC optic atrophy, deafness, sensitivity to sunlight, and mental  
CC retardation.  
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.  
CC -1- DATABASE: NAME-Atlas Genet. Cytogenet. Oncol. Haematol.;  
CC WWW="http://www.infobiogen.fr/services/chromocancer/genes/CSBID302.html".  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: L04791; AAA52397.1; -;  
DR PIR: A44224; A44224;  
DR TRANSFAC: T04550; -;  
DR Genew: HGNC:3438; ERCC6.  
DR MIM: 133540; -;  
DR InterPro: IPR001410; DEAD.  
DR InterPro: IPR001650; Helicase\_C.  
DR InterPro: IPR000330; SNF2\_N.  
DR Pfam: PF00176; SNF2\_N.1.  
DR Pfam: PF00271; Helicase\_C.1.  
DR SMART: SM00487; DEXDC; 1.  
DR SMART: SM00490; HELIC; 1.  
DR Helicase; DNA repair; ATP-binding; DNA-binding; Nuclear protein;  
KW Transcription regulation; Cockayne's syndrome; Disease mutation;  
KW Polymorphism.  
FT DOMAIN 356 394 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 446 446 GLY-RICH.  
FT DOMAIN 466 481 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT DOMAIN 1038 1055 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT NP\_BIND 532 539 ATP (POTENTIAL).  
FT SITE 646 649 DECH BOX.  
FT VARIANT 255 255 K -> T.  
FT VARIANT 399 399 /FTID=VAR\_001216.  
FT VARIANT 670 670 G -> D.  
FT VARIANT 670 670 /FTID=VAR\_001217.  
FT VARIANT 851 851 R -> W (IN CSB).  
FT VARIANT 851 851 /FTID=VAR\_001218.  
FT VARIANT 957 957 W -> R (IN CSB).  
FT VARIANT 957 957 /FTID=VAR\_001219.  
FT VARIANT 1042 1042 V -> G (IN CSB).  
FT VARIANT 1042 1042 /FTID=VAR\_001220.  
FT VARIANT 1095 1095 P -> L (IN CSB).  
FT VARIANT 1095 1095 /FTID=VAR\_001221.  
FT VARIANT 1097 1097 P -> R (IN CSB).  
FT VARIANT 1097 1097 /FTID=VAR\_001222.  
FT VARIANT 1213 1213 M -> V.  
FT VARIANT 1213 1213 /FTID=VAR\_001223.  
FT VARIANT 1413 1413 R -> G (IN CSB).  
FT VARIANT 1413 1413 /FTID=VAR\_001224.  
FT VARIANT 1413 1413 O -> R.  
FT VARIANT 1413 1413 /FTID=VAR\_001225.  
SO SEQUENCE 1493 AA: 168415 MM: 285257E2AEC071AC CRC64;  
Query Match 7.0%; Score 133; DB 1; Length 1493;

Best Local Similarity 23.7%; Pred. No. 1.9;  
Matches 82; Conservative 42; Mismatches 118; Indels 104; Gaps 17;  
QY 35 GAERVP-----EEDSRMSRAFPOLGCRPGEGSGLESOPPPLOTQACPESSCLRGG 88  
DB 201 GAERVKIELDHLASLEDA-----EPGP---SLGSLMIPVOETAWME--LIRIG 243  
QY 89 EKQMGDDSSAGDFPPPAVEPPTP---EA-----ELLAQPCDHSEASK 129  
DB 244 QM-----TPFGQITQKQEKRRKIMLNASFEKYLADQALTSERKKQGNKRAAK 297  
QY 130 LGAPA-----AGGEEMGQOOROLGKKHRRSPKRRHMKPYKLTWEKKRDEKQSL 184  
DB 298 APAPVPPAPVQKNKPNKRAVLSKKEE-----LKHKKLQKRALQFGKVGLPKAR 352  
QY 185 RA-----SRIRAEPAFGQAPAPYNTQFLMDHDQ-----EPDL-----KRG 223  
DB 353 RPWESDMRPAEEDSGEE-SEYPTEEEEEDDEVEGAADLSDGTDYELKPLPKG 411  
QY 224 LYSKRAAKSDOTSDDPFMEEGEE-DGSGDGGGSGSEFLQDPFSEYERYHTESLQNN 282  
DB 412 KRQKKVPVGE---IDDDFFSSGEELAAASVGGGGGGRKVGRRDGDDEYKQRLRW 468  
QY 283 SKQELIKEYLELEKCLSMEDENNRLRLSKRLGDDARVLELE 328  
DB 469 NK-----LRLQDKERKRLKLE-----DDSESDAEFD 494  
RESULT 8  
TRHY\_HUMAN  
ID TRHY\_HUMAN STANDARD; PRT; 1898 AA.  
AC 007283;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Trichohyalin.  
GN THH OR TRHY OR THL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_Taxid:9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93280194; PubMed=7685034;  
RA Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D.,  
RA Steinhert P.M.;  
RT "The structure of human trichohyalin. Potential multiple roles as a  
RT functional EF-hand-like calcium-binding protein, a cornified cell  
RT envelope precursor, and an intermediate filament-associated (cross-  
RT linking) protein.";  
RT J. Biol. Chem. 268:12164-12176(1993).  
RN [2]  
RP SEQUENCE OF 1731-1898 FROM N.A. AND CHARACTERIZATION.  
RX MEDLINE=93315897; PubMed=7686933;  
RA O'Keefe E.J., Hamilton E.H., Lee S.-C., Steinhert P.M.;  
RA "Trichohyalin: a structural protein of hair, tongue, nail, and  
RA epidermis.";  
RL J. Invest. Dermatol. 101:655-715(1993).  
CC -1- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES  
CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE  
CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR  
CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY  
CC ISOPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER  
CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN  
CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN  
CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL  
CC DIFFERENTIATION.  
CC -1- SUBUNIT: MONOMER (PROBABLE).  
CC -1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS  
CC THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA AND IN  
CC THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).  
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF  
CC THE EPIDERMIS.



```

CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRAINED
CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
CC DIFFERENT SPECIES.
CC -1- PTM: SUBSTRATE OF TRANSGUTAMINASE. SOME 200 ARGININES ARE
CC PROBABLY CONVERTED TO CITULLINES BY PEPTIDYLARGININE DEIMIDASE.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE S-100
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: L09190; AAA65582.1; -.
CC DR PIR: A45973; A45973.
CC DR HSSP: P02633; A1CB.
CC DR Genew: HGNC:11791; THH.
CC DR MIM: 190370; -.
CC DR InterPro: IPR001751; CAP_S100.
CC DR InterPro: IPR002048; EF-hand.
CC DR InterPro: IPR002017; Spectrin.
CC DR Pfam: PF00036; ehfand; 1.
CC DR ProDom: PD003407; CAP_S100; 1.
CC DR ProSITE: PS00018; EF_HAND; 1.
CC DR ProSITE: PS00303; S100_CAP; 1.
CC KM Keratinization; Repeat; Calcium-binding.
CC FT DOMAIN 1
CC FT CA_BIND 22 33
CC FT CA_BIND 62 73
CC FT DOMAIN 314 390
CC FT REPEAT 314 326
CC FT REPEAT 327 339
CC FT REPEAT 340 351
CC FT REPEAT 352 364
CC FT REPEAT 365 377
CC FT REPEAT 378 390
CC FT DOMAIN 391 444
CC FT REPEAT 391 396
CC FT REPEAT 397 402
CC FT REPEAT 403 408
CC FT REPEAT 409 414
CC FT REPEAT 415 420
CC FT REPEAT 421 426
CC FT REPEAT 427 432
CC FT REPEAT 433 438
CC FT REPEAT 439 444
CC FT DOMAIN 444 702
CC FT REPEAT 444 444
CC FT DOMAIN 923 1162
CC FT REPEAT 923 952
CC FT REPEAT 953 982
CC FT REPEAT 983 1012
CC FT REPEAT 1013 1042
CC FT REPEAT 1043 1072
CC FT REPEAT 1073 1102
CC FT REPEAT 1103 1132
CC FT REPEAT 1133 1162
CC FT DOMAIN 1250 1849
CC FT REPEAT 1250 1752
CC FT REPEAT 1752 1801
CC FT REPEAT 1801 1857
CC FT REPEAT 1857 1880
CC FT CONFLICT 1880

```

Seq	SEQUENCE	1898 AA:	247219 MM:	A7AB5947FEB62E3ID CRC64:
S0	SEQUENCE	1898 AA:	247219 MM:	A7AB5947FEB62E3ID CRC64:
	Query Match	6.9%	Score 131:	DB 1: Length 1898:
	Best Local Similarity	22.4%	Pred. No. 3.2:	
	Matches	79: Conservative	54: Mismatches	146: Indels 74: Gaps 13:
OY	31 ERPGAER-----VPEDSRMOSRAFPOLGCRPGPEGCSLESOPPLQTOACRESS	83		
DB	813 EROIAAEERQOREORFLPREEEKFOR-----GRORRERKEKLQPLEEEOLORRERAQ	865		
OY	84 CLRECEGQNDSDSAGGDPFPPAEVEPTPAE-----LLAOPCHDSASKIGAPAG	136		
DB	866 QLOEEEDGLOEDQERRROEQRRDQKRWQLEEEKRRRHTIYAKPALQEOLRKQDLOE	925		
OY	137 GEEWGOOOROLGKKHRRRPSKKRMKPYKLTWEKKKFDKOSLRASRTAEMFAK	196		
DB	926 EEELQOEERERE-----KRRQOEEROYREEEOLOQEEOLLRERREKRRROEEROYRK	979		
OY	197 GQVAPYNTTQFLMDHDQ---EEPDILKTGLYSKRAAKSDOTSDDPFMEBGEEEDGCS	253		
DB	980 DKR-----LQKEEEOGLGEEPEKR-----RROEKKKYREEEEOE-----	1016		
OY	254 GMGDSGSEPLQDRDSEIYER-----YHTESLQMSKOELIKETLEBKLSRMD--FN	305		
DB	1017 -----EOILLRERERKRROEMEROYRKDELOOEEOLLRERERKRRLQEROERYEE	1070		
OY	306 NRLLESKRLGDDAVR---ELELEL---DLRAENLQLTEN-ELHROER	351		
DB	1071 EELQEEEOGLGEEERTRRROELEORYRKDELOOEEOLLRERERKRROER	1123		
RESULT 9				
CMGA-RAT				
ID	CMGA-RAT	STANDARD:	PRT:	466 AA.
AC	PI0354:			
DT	01-MAR-1989 (Rel. 10, Created)			
DT	01-MAR-1989 (Rel. 10, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Chromogranin A precursor (CGA) [contains: Pancreastatin; Beta-granin; WE-141].			
DE	CHGA.			
GN	Rattus norvegicus (Rat).			
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_Taxid=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=8611232; Pubmed=2828116;			
RA	Iacangelo A., Okayama H., Eiden L.E.;			
RT	"Primary structure of rat chromogranin A and distribution of its mRNA";			
RT	FEBs Lett. 227:115-121(1988).			
RL	[3]			
RN	SEQUENCE OF 19-32.			
RP	MEDLINE=85285598; Pubmed=3896848;			
RX	Hutton J.C., Hansen F., Peshavaria M.;			
RA	"Beta-granins: 21 kDa co-secreted peptides of the insulin granule closely related to adrenal medullary chromogranin A";			
RL	FEBs Lett. 188:336-340(1985).			
CC	-1- FUNCTION: PANCREASTATIN STRONGLY INHIBITS GLUCOSE INDUCED INSULIN RELEASE FROM THE PANCREAS.			
CC	-1- SUBCELLULAR LOCATION: Neuroendocrine and endocrine secretory granules.			
CC	-1- PTM: CGA IS O-GLYCOSYLATED.			
CC	-1- MISCELLANEOUS: BINDS CALCIUM WITH A LOW-AFFINITY.			



```
KW Chromosomal translocation; Proto-oncogene; Alternative splicing.
FT DOMAIN 36 206 RAS-INTERACTING.
FT DOMAIN 425 491 FHA.
FT DOMAIN 804 910 DILUTE.
FT DOMAIN 991 1077 PDZ.
FT DOMAIN 162 174 GLU/LYS-RICH.
FT DOMAIN 1349 1356 POLY-PRO.
FT DOMAIN 1371 1376 POLY-PRO.
FT DOMAIN 1561 1571 ASP/GLU-RICH (ACIDIC).
FT SITE 26 26 ML FUSION POINT (IN AN ACUTE MYELOID
LEUKEMIA PATIENT).
FT VARSPLIC 1588 1611 LODERRQQLLEMRREADRA -> VKGVLMLCESV
FT VARSPLIC 1612 1816 PILASCEPMG (IN ISOFORM 1).
FT VARSPLIC 1666 1743 MISSING (IN ISOFORM 1).
FT LCRPPLPDYPPSPAPGAPPPPPORPMASLYLKTQVLPD
FT SLTAKFYANNEEEEDCCSIAGODKSSMPKSHGL ->
FT PNSTPGSTGAAGVAGHADACROKRRKSSQDDSGSSCAP
FT NLTKRORLFSQGDVSNKVKASRKLTLELELNTR (IN
FT ISOFORM 3).
FT VARSPLIC 1744 1816 MISSING (IN ISOFORM 3).
FT CONFLICT 373 373 G -> V (IN REF. 1).
FT CONFLICT 391 391 P -> PGRNHFAVYNTHTYE (IN REF. 3).
FT CONFLICT 744 744 D -> DSHED (IN REF. 3).
FT CONFLICT 1031 1031 D -> DV (IN REF. 1).
FT CONFLICT 1408 1408 R -> P (IN REF. 1).
SQ SEQUENCE 1816 AA; 205604 MW; EBIJFEF04879CE8F CRC64;

Query Match 6.8%; Score 129.5; DB 1; Length 1816;
Best Local Similarity 20.3%; Pred. No. 3.7;
Matches 78; Conservative 51; Mismatches 144; Indels 111; Gaps 14;

QY 14 QTSNCTGAAYVOEELN--PERPPGAEEKVRPEEDSRMOS-RAFPOLGGRPGEGGLESQ 70
DB 1291 QSSLSIDSTSSQEHNLNSSKSVTPASTLTSGPRMKTPTAIPATPAVAVSQPIRTDLPP 1350
QY 71 PPPIQTOACPESSCLREGKCGNGDDSSAGDPPRPPEVEPTPAELLLOPCHDSEASKL 130
DB 1351 PPPPPVHYA-----GDFGMSMDLPLP---PPP-----SANQI 1380
QY 131 GAPPAAGGEEMWGOOQROLGKKKRRRPSKKRHKPYUKLWE-EKKKFKDEKOSLRASRI 189
DB 1381 GLPFA-----OVAAEKRRKREHONWYKEKARLEEEERKRRERLQGMRT 1429
QY 190 RAEMFANGQVPARYNTQF-----LMDHDQEEP-----DLKTGLYSKR 228
DB 1430 QS---LNPAPFSPPLTAQOMKPEKSTLQRPQETVIRELQPOQPRITERRDLQYITYSKE 1486
QY 229 AAASDSDTSDDEMEEGEGEDGSGDMGCGSEFLQRPDFTTYRYHTESLOJNSKOFELI 288
DB 1487 ELSSGDSLSPDPW-----KRDAKKLEKQOQMHTVDM-----LS 1520
QY 289 KEYLELEKCLSRMEDENNRLR-----LESKRLLGGDDARVRELE-----LDRL 332
DB 1521 KEIPELOSKPRPSAEESRLKLMLEWFOFKRLOESKOKDEDEEEDDDVDTMLIMORL 1580
QY 333 RAENIQLTENELHRQOERAPLSK 356
DB 1581 EAERRARLODEERRRQQLLEMRK 1604

RESULT 11
VNUA_VNUA STANDARD: PRT; 1733 AA.
AC P33485;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE Probable nuclear antigen.
OS Pseudorabies virus (strain Kaplan) (PRV).
OC Viruses; dsDNA viruses, no RNA stages; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirinae.
NCBI_TaxID=33703;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=91021039; Pubmed=2171211;
RA Vilek C.; Kozmik Z.; Paces V.; Schirm S.; Schwyzer M.;
RT "Pseudorabies virus immediate-early gene overlaps with an oppositely
RT oriented open reading frame: characterization of their promoter and
RT enhancer regions.";
RL Virology 179:365-377(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M34651; AAA47471.1; -.
DR PIR; B45344; B45344.
FT DOMAIN 112 117 POLY-THR.
FT DOMAIN 179 1733 GLY-RICH.
FT DOMAIN 192 196 POLY-SER.
FT DOMAIN 271 298 POLY-PRO.
FT DOMAIN 304 308 POLY-ARG.
FT DOMAIN 883 889 POLY-GLY.
FT DOMAIN 1398 1405 POLY-GLY.
SQ SEQUENCE 1733 AA; 172166 MW; 0C8CD8BE475BB5E2 CRC64;

Query Match 6.7%; Score 128; DB 1; Length 1733;
Best Local Similarity 22.4%; Pred. No. 4.3;
Matches 74; Conservative 27; Mismatches 94; Indels 136; Gaps 16;

QY 8 EYOHQPTSNCT-----GAAAYOEE-----LNPER-PPGAEEK----- 39
DB 104 DHQHRPPPTTTTITKQPHQBDLRLPYTLQDEEDHLLRPTYPDPSSAKTNNHQDPGCG 163
QY 40 VPEEDSRMOSRAFPOLGGRPP-----EGEGSLESQPPPLQ----- 75
DB 164 GPRSTSSHHNHQDPGCGPRPPRPSTSSSHQGPSTRRPPPOPPRPWPPPPSPQKI 223
QY 76 --TQACESSCLR-----EGEKQNGDDSSAGG-----DPPPAEVEPIPEA 115
DB 224 SETPAGSENNATQTLFSSHSEKLLFSPMGEGEG-GRGTAGCGEGEDRDPPPPSPPPPP 282
QY 116 ELIAQPCDHSEASKLGAAPAGGEEMWGOOQROLG---KKKRRRPSKKRHKPYUKLWTW 172
DB 283 PLPPPPPPPPPPQ---PPPAAGS---ARRRRGGGPPGRCGRKRGKRRR----- 326
QY 173 EEKKKFKDEKOSLRASIRAEFAKQVPARYNTQFLMDHDQEEPDLKTGLYSKRAAAK 232
DB 327 -----ACGTFAA-----AADAEEFEDGD----- 344
QY 233 SDOISDDDFMEEGEGEDGSGD--GMGDCGSE 261
DB 345 GDEDEDEDRAEGEGEGEDGEGPRGAGGAGE 375

RESULT 12
ATRX_MOUSE
ID ATRX_MOUSE STANDARD: PRT; 2476 AA.
AC Q61687;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcriptional regulator ATRX (X-linked nuclear protein)
DE (heterochromatin protein 2) (HPI alpha-interacting protein) (HPI-BP38
DE protein).
DE ATRX OR XNP OR HPIBP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
```

RX MEDLINE=98213653; PubMed=9545503;  
 RA Picketts D.J., Tastan A.O., Higgs D.R., Gibbons R.J.;  
 RT "Comparison of the human and murine ATRX gene identifies highly  
 RT conserved, functionally important domains";  
 RL Mamm. Genome 9:400-403(1998).  
 RN [2]  
 RP SEQUENCE OF 325-1176 FROM N.A.  
 RX MEDLINE=97133299; PubMed=8978696;  
 RA le Douarin B., Nielsen A.L., Garnier J.-M., Ichinose H.,  
 RA Jeanmougin F., Losson R., Chambon P.;  
 RT "A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic  
 RT control of transcription by nuclear receptors";  
 RL Embo J. 15:6701-6715(1996).  
 RN [3]  
 RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC  
 RP HETEROCHROMATIN.  
 RX MEDLINE=20040663; PubMed=10570185;  
 RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,  
 RA Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,  
 RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;  
 RT "Localization of a putative transcriptional regulator (ATRX) at  
 RT pericentromeric heterochromatin and the short arms of acrocentric  
 RT chromosomes";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).  
 CC -1- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES  
 CC GENE EXPRESSION BY AFFECTING CHROMATIN.  
 CC -1- SUBUNIT: PROBABLY BINDS EZH2. BINDS ANNEXIN V IN A CALCIUM AND  
 CC PHOSPHATIDYLCHOLINE/PHOSPHATIDYLSELINE-DEPENDENT MANNER (BY  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC  
 CC HETEROCHROMATIN DURING INTERPHASE AND MITOSIS, PROBABLY BY  
 CC INTERACTING WITH HP1.  
 CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AF026032; AAC08741.1; -;  
 DR EMBL: X99643; CA67962.1; -;  
 DR MGD: MGI:103067; Xnp.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR000330; SNF2\_N.  
 DR InterPro: IPR001841; ZnF\_Finger.  
 DR Pfam: PF00176; SNF2\_N; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR SMART: SM00490; HELIC; 1.  
 DR SMART: SM00184; RING; 1.  
 KM DNA repair: Nuclear protein; DNA-binding; Helicase; ATP-binding;  
 KM Zinc-finger.  
 FT ZN\_FING 219 267 PHD-TYPE.  
 FT NP\_BIND 1579 1586 ATP (POTENTIAL).  
 FT SITE 1704 1707 DEGH BOX.  
 FT DOMAIN 319 322 POLY-SER.  
 FT DOMAIN 735 738 POLY-SER.  
 FT DOMAIN 1001 1004 POLY-GLU.  
 FT DOMAIN 1130 1135 POLY-SER.  
 FT DOMAIN 1182 1185 POLY-SER.  
 FT DOMAIN 1238 1245 POLY-ASP.  
 FT DOMAIN 1484 1487 POLY-GLU.  
 FT DOMAIN 1924 1931 POLY-SER.  
 FT DOMAIN 2205 2208 POLY-LYS.  
 FT DOMAIN 2248 2248 POLY-GLU.  
 FT DOMAIN 2403 2408 POLY-GLN.  
 SO SEQUENCE 2476 AA; 278601 MW; 90A42B790FC4FF4C CRC64;

Query Match 6.7%; Score 128; DB 1: Length 2476;  
 Best Local Similarity 19.7%; Pred. No. 6.2;  
 Matches 78; Conservative 52; Mismatches 130; Indels 136; Gaps 13;  
 QY 43 EDSRQSRAPFQOLGPEPEGE---GSLSEOPPLQTOACPESSCLRGEGGQNGDDSSA 99  
 DB 1142 EDNRKQKQRTSARKKTGNTKEKRNSLRATPKROYDITSSSPDGDDNSAGSESSD 1201  
 QY 100 GGDPPP-----PAVEP-----TPE-----AELLA 119  
 DB 1202 EOKIKPTVENLVDSHTGFCQSSGDALSKSVPAIVDDDDNDPNNRIAKMLLEETKA 1261  
 QY 120 QPCHDSEASKLGAPAAAGEEEMGQOQ-----KQAGKKHRRRPSKK- 161  
 DB 1262 NLSSDEGSSDDEPDGCGKKRTGKQSESPADDELREQLAVQNVSESSDSEESKRP 1321  
 QY 162 --RHKKPYKLTWEKKKKFDEKQSLRASRIAEAFKQAPVAPYNTTQFLMDHDQDEPD 219  
 DB 1322 RYRHLRLHKKLTLSSGESEGEK---PKPREHREAKGR----- 1356  
 QY 220 LKTGLYSKRAAKSDPTSDDFMEGGEEDGSGMGSGSEFLQRPSETERYHTESL 279  
 DB 1357 -----NRRKVSSEDSDDTFOESGVSEVSE-----SEDRQRPRTSRAKAELEEN 1402  
 QY 280 QMSKQELIKELYELEKCLR-----MEDENNRLRESKRL 315  
 DB 1403 QRSYKQKKRRRRIRVQEDSSSEKSHSEDEKKEGDEDEDEDEDEDEND---DSKSP 1458  
 QY 316 GGDARVRELELDRLAEMLQLTTELHROER 351  
 DB 1459 GKGRKKIRKI-LKDDKILTE-----TONALKKEER 1488  
 RESULT 13  
 ID PLEL\_RAT STANDARD; PRT; 4687 AA.  
 AC P30427; O08879; O08880; O08881;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Plectin 1 (PLTN) (PCN).  
 GN Plect.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Glial tumor;  
 RX MEDLINE=91268156; PubMed=2050743;  
 RA Wiche G., Becker B., Lubert K., Weitzer G., Castanon M.J.,  
 RA Hauptmann R., Stralowa C., Stewart M.;  
 RT "Cloning and sequencing of rat plectin indicates a 466-kD polypeptide  
 RT chain with a three-domain structure based on a central alpha-helical  
 RT coiled coil";  
 RL J. Cell Biol. 114:83-99(1991).  
 RN [2]  
 RP REVISIONS.  
 RC TISSUE=Glial tumor;  
 RX MEDLINE=96210632; PubMed=8633055;  
 RA Liu C.-G., Maercker C., Castanon M.J., Hauptmann R., Wiche G.;  
 RT "Human plectin: Organization of the gene, sequence analysis, and  
 RT chromosome localization (8q24).";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283(1996).  
 RN [3]  
 RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4), AND TISSUE SPECIFICITY.  
 RC TISSUE=Glial tumor;  
 RX MEDLINE=97321050; PubMed=9177781;  
 RA Elliott C.E., Becker B., Oehler S., Castanon M.J., Hauptmann R.,  
 RA Wiche G.;  
 RT "Plectin transcript diversity: identification and tissue distribution  
 RT of variants with distinct first coding exons and rodless isoforms";  
 RL Genomics 42:115-125(1997).

CC -1- FUNCTION: INTERLINKS INTERMEDIATE FILAMENTS WITH MICROTUBULES AND  
 CC MICROFILAMENTS AND ANCHORS INTERMEDIATE FILAMENTS TO DESMOSOMES OR  
 CC HEMIDESMOSOMES. MAY BE INVOLVED NOT ONLY IN THE CROSSLINKING AND  
 CC STABILIZATION OF CYTOSKELETAL INTERMEDIATE FILAMENTS NETWORK, BUT  
 CC ALSO IN THE REGULATION OF THEIR DYNAMICS.  
 CC -1- SUBUNIT: HOMODIMER OR HOMOTETRAMER.  
 CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS; 1 (SHOWN HERE), 2, 3 AND 4; ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST EXPRESSION IN  
 CC SKELETAL MUSCLE AND LOWEST IN THYMUS.  
 CC -1- DOMAIN: THE N-TERMINUS INTERACTS WITH ACTIN, THE C-TERMINUS WITH  
 CC VIMENTIN, DESMIN, GFAP, CYTOKERATINS, LAMIN B; WHEREAS BOTH THE N-  
 CC AND THE C-TERMINUS CAN BIND INTEGRIN BETA-4.  
 CC -1- PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM  
 CC INTERMEDIATE FILAMENTS DURING MITOSIS (BY SIMILARITY).  
 CC -1- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 3 PLECTIN REPEATS.  
 CC -1- SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.  
 CC -1- SIMILARITY: BELONGS TO THE PLEKIN OR CYTOLINKER FAMILY.  
 CC -1- CAUTION: ISOFORM 4 IS A FRAGMENT AT THE N-TERMINUS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb.sib.ch](mailto:license@isb.sib.ch)).  
 CC -----  
 DR EMBL; X59601; CAA42169.1; -;  
 DR EMBL; U96274; AAC53209.1; -;  
 DR EMBL; U96275; AAC53210.1; -;  
 DR EMBL; U96276; AAC53211.1; -;  
 DR PIR; A39638; A39638.  
 DR PIR; S21876; S21876.  
 DR HSSP; Q01082; 1BKR.  
 DR InterPro: IPR001589; Actbind\_actnin.  
 DR InterPro: IPR001715; Calponin-like.  
 DR InterPro: IPR001101; Plectin\_repeat.  
 DR InterPro: IPR005326; S10\_Plectin.  
 DR InterPro: IPR002017; Spectrin.  
 DR Pfam; PF00307; CH; 2.  
 DR Pfam; PF00681; Plectin; 21.  
 DR Pfam; PF03501; S10\_Plectin; 1.  
 DR SMART; SM00033; CH; 2.  
 DR SMART; SM00250; PLEC; 33.  
 DR SMART; SM00150; SPEC; 4.  
 DR PROSITE; PS00019; ACTININ\_1; FALSE\_NEG.  
 DR PROSITE; PS00020; ACTININ\_2; FALSE\_NEG.  
 DR PROSITE; PS50021; CH; 2.  
 KW Coiled coil: Repeat: Structural protein: Cytoskeleton: Actin-binding;  
 KW Phosphorylation; Alternative splicing.  
 FT DOMAIN 1 1473 GLOBULAR 1.  
 FT DOMAIN 1474 2758 CENTRAL FIBROUS ROD DOMAIN.  
 FT DOMAIN 2759 4687 GLOBULAR 2.  
 FT DOMAIN 181 406 ACTIN-BINDING.  
 FT DOMAIN 185 288 CH 1.  
 FT DOMAIN 301 403 CH 2.  
 FT DOMAIN 648 722 SPECTRIN 1.  
 FT REPEAT 743 827 SPECTRIN 2.  
 FT REPEAT 840 933 SPECTRIN 3.  
 FT REPEAT 1318 1418 SPECTRIN 4.  
 FT DOMAIN 1472 1692 COILED COIL (POTENTIAL).  
 FT DOMAIN 1724 2760 COILED COIL (POTENTIAL).  
 FT REPEAT 2791 2828 PLECTIN 1.  
 FT REPEAT 2829 2866 PLECTIN 2.  
 FT REPEAT 2867 2904 PLECTIN 3.  
 FT REPEAT 2905 2942 PLECTIN 4.  
 FT REPEAT 2943 2980 PLECTIN 5.  
 FT REPEAT 2984 3018 PLECTIN 6.  
 FT REPEAT 3119 3156 PLECTIN 7.  
 FT REPEAT 3157 3194 PLECTIN 8.

FT REPEAT 3195 3232 PLECTIN 9.  
 FT REPEAT 3233 3270 PLECTIN 10.  
 FT REPEAT 3271 3308 PLECTIN 11.  
 FT REPEAT 3311 3346 PLECTIN 12.  
 FT REPEAT 3348 3385 PLECTIN 13.  
 FT REPEAT 3526 3563 PLECTIN 14.  
 FT REPEAT 3564 3601 PLECTIN 15.  
 FT REPEAT 3602 3639 PLECTIN 16.  
 FT REPEAT 3643 3677 PLECTIN 17.  
 FT REPEAT 3823 3860 PLECTIN 18.  
 FT REPEAT 3861 3898 PLECTIN 19.  
 FT REPEAT 3899 3936 PLECTIN 20.  
 FT REPEAT 3937 3974 PLECTIN 21.  
 FT REPEAT 3978 4011 PLECTIN 22.  
 FT REPEAT 4066 4103 PLECTIN 23.  
 FT REPEAT 4104 4141 PLECTIN 24.  
 FT REPEAT 4142 4179 PLECTIN 25.  
 FT REPEAT 4180 4217 PLECTIN 26.  
 FT REPEAT 4221 4255 PLECTIN 27.  
 FT REPEAT 4268 4308 PLECTIN 28.  
 FT REPEAT 4411 4448 PLECTIN 29.  
 FT REPEAT 4449 4486 PLECTIN 30.  
 FT REPEAT 4487 4524 PLECTIN 31.  
 FT REPEAT 4525 4562 PLECTIN 32.  
 FT REPEAT 4563 4600 PLECTIN 33.  
 FT DOMAIN 4253 4303 BINDING TO INTERMEDIATE FILAMENTS.  
 FT DOMAIN 4628 4643 4 X 4 AA TANDEM REPEATS OF G-S-R-X.  
 FT MOD\_RES 4542 4542 PHOSPHORYLATION (BY CDC2) (BY  
 FT SIMILARITY).  
 FT VARSPLIC 1 180  
 FT EMBL; X59601; CAA42169.1; -;  
 FT EMBL; U96274; AAC53209.1; -;  
 FT EMBL; U96275; AAC53210.1; -;  
 FT EMBL; U96276; AAC53211.1; -;  
 FT PIR; A39638; A39638.  
 FT PIR; S21876; S21876.  
 FT HSSP; Q01082; 1BKR.  
 FT InterPro: IPR001589; Actbind\_actnin.  
 FT InterPro: IPR001715; Calponin-like.  
 FT InterPro: IPR001101; Plectin\_repeat.  
 FT InterPro: IPR005326; S10\_Plectin.  
 FT InterPro: IPR002017; Spectrin.  
 FT Pfam; PF00307; CH; 2.  
 FT Pfam; PF00681; Plectin; 21.  
 FT Pfam; PF03501; S10\_Plectin; 1.  
 FT SMART; SM00033; CH; 2.  
 FT SMART; SM00250; PLEC; 33.  
 FT SMART; SM00150; SPEC; 4.  
 FT PROSITE; PS00019; ACTININ\_1; FALSE\_NEG.  
 FT PROSITE; PS00020; ACTININ\_2; FALSE\_NEG.  
 FT PROSITE; PS50021; CH; 2.  
 KW Coiled coil: Repeat: Structural protein: Cytoskeleton: Actin-binding;  
 KW Phosphorylation; Alternative splicing.  
 FT DOMAIN 1 1473 GLOBULAR 1.  
 FT DOMAIN 1474 2758 CENTRAL FIBROUS ROD DOMAIN.  
 FT DOMAIN 2759 4687 GLOBULAR 2.  
 FT DOMAIN 181 406 ACTIN-BINDING.  
 FT DOMAIN 185 288 CH 1.  
 FT DOMAIN 301 403 CH 2.  
 FT DOMAIN 648 722 SPECTRIN 1.  
 FT REPEAT 743 827 SPECTRIN 2.  
 FT REPEAT 840 933 SPECTRIN 3.  
 FT REPEAT 1318 1418 SPECTRIN 4.  
 FT DOMAIN 1472 1692 COILED COIL (POTENTIAL).  
 FT DOMAIN 1724 2760 COILED COIL (POTENTIAL).  
 FT REPEAT 2791 2828 PLECTIN 1.  
 FT REPEAT 2829 2866 PLECTIN 2.  
 FT REPEAT 2867 2904 PLECTIN 3.  
 FT REPEAT 2905 2942 PLECTIN 4.  
 FT REPEAT 2943 2980 PLECTIN 5.  
 FT REPEAT 2984 3018 PLECTIN 6.  
 FT REPEAT 3119 3156 PLECTIN 7.  
 FT REPEAT 3157 3194 PLECTIN 8.

Query Match 6.7% Score 127.5; DB 1; Length 4687;  
 Best Local Similarity 21.4%; Pred. No. 13;  
 Matches 78; Conservative 51; Mismatches 127; Indels 109; Gaps 13;  
 25 QEELNRPFGAEERPEEDSRWOSRAPFOLGPGPEEGSLSESOPPLIQOACPSSC 84  
 1755 QQQAELER---AREAEARELERWQLKANELRLR-----LQAEVAQQS 1796  
 85 LRPGEKGQNDSSAGGDFPPPAEVEPTPAELLAPCH-----DSEAS 128  
 1797 LAQADAEKQKEAEAREARRRKAEDQAVRQLEAEQELKOROLTEGTAQOQRIAAEOEL 1856  
 129 KLGAPAAAGGEEMGOOOR-----LQKKHRRRPSKKKRWKYYLTMEKKKPFERKS 183  
 1857 RLRA-----ETDEGEHQRQLEBEELARLDQENATATQKRO-----ELEELAVRAEMEV 1906  
 184 LRASRTAEWFAKQGVVADYNTTQFLMDHDOEPDLTKGLYSKRA-----AA 231  
 1907 LLSKARAEDESS-----TSKSKQRLEAEAGRRRELAEEAARLRALAEAR 1954  
 232 KSDDTDDDPMEGGEGSDGSGMGDSEFLQDRPSEYERYHTESLQNMKSQELIKET 291  
 1955 RHRELAEEPAARQRAEADG-----VLTEKLAIAISKATRLK-- 1989

QY 292 LELEKLSRMEDENNRLLESKRIGDDA-RVLELELDRLRAE-----NQLLTENE 344  
 DB 1990 TEAIALKEAEKAEERLERL-----RLAEDEAFORRLLEQAOAHQADIEERLALQLRKASESE 2045  
 QY 345 LHRQO 349  
 DB 2046 LEROK 2050

RESULT 14  
 TRHY\_RABIT  
 ID TRHY\_RABIT STANDARD: PRT: 1407 AA.  
 AC P37709;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, last sequence update)  
 DT 15-JUN-2002 (Rel. 41, last annotation update)  
 DE Trichohyalin.  
 GN THH.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Fietz M.J., Rogers G.E.;  
 RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.  
 CC FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES  
 IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE  
 INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR  
 LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY  
 ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER  
 WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN  
 ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN  
 ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL  
 DIFFERENTIATION.  
 CC SUBUNIT: HOMODIMER (PROBABLE).  
 CC TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS  
 THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN  
 THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).  
 CC DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF  
 THE EPIDERMIS.  
 CC DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND  
 CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST  
 ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS  
 OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED  
 ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS  
 THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.  
 CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN  
 THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG  
 DIFFERENT SPECIES.  
 CC PTM: SUBSTRATE OF TRANSGUTAMINASE. SOME 200 ARGININES ARE  
 PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.  
 CC SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100  
 FAMILY  
 CC SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: Z19092; CAA79519.1;  
 DR PIR: S28589; S28589.  
 DR HSP: P02633; 41CB.  
 DR InterPro: IPR001751; CAP\_S100.  
 DR InterPro: IPR002048; EF-hand.  
 DR Pfam: PF00036; efhand.1.  
 DR Pfam: PF01023; S\_100.1.  
 DR ProDom: PD003407; Cap\_S100.1.

DR PROSITE: PS00018; EF\_HAND.1.  
 DR PROSITE: PS00303; S100\_CAP.1.  
 KW Keratinization; Repeat; Calcium-binding.  
 FT DOMAIN 1 91 S-100 Like.  
 FT CA\_BIND 22 33 EF\_HAND 1 (LOW AFFINITY) (POTENTIAL).  
 FT CA\_BIND 62 73 EF\_HAND 2 (HIGH AFFINITY) (POTENTIAL).  
 SO SEQUENCE 1407 AA: 183781 MW: AET7DZAI59F12BVF CXC64;  
 Query Match  
 Best local Similarity 23.7%; Score 126.5; DB 1; Length 1407;  
 Matches 83; Conservative 48; Mismatches 124; Indels 95; Gaps 17;  
 QY 25 QEE-----INPERPGAEEVPEEDSRWOSRAFPLQGRPGEGSLESQPPPIQTOAC 79  
 DB 349 OEERRQLLAEEVREQARGESILTRMQLQESFAGAR-----QSKVSRPRQEGES- 402  
 QY 80 PESSCLREGKCGNGDSSAGSDPFPFAVEVPPPEVLAQPCHDSEAKLGAAPAGGE- 139  
 DB 403 -----LRQDERROER-----ERELEQ-----ARRQO 427  
 QY 140 EW-GGQOROLGKKHRRPSKKRHKPKYKLTWEKKRPDEKOSLRASRAEMPAKQ 198  
 DB 428 QMGAESEERRRRLSARPSLRERQLRAER--QEQQRREEDQRER-RQL----- 479  
 QY 199 PVAPYNTQFLMDHDQEEPLDKTGLYSKRAAKSDTSDDEMEEGEGSDGMDGD 258  
 DB 480 -----GFL-----EEFQL-----QRERAQQLQEDSDQEDRERRRRQEQRPQO 520  
 QY 259 GSEFLQDFSEYERYHT-----ESLQ-NMSKQLLIEYLEKLSRMD 303  
 DB 521 TWRWQLQE--EAQRRTTLYAKPGQEOLEREBELQREKRQREERYEKEE-LQRED 577  
 QY 304 ENNRLLSKRLGDDARVARELELDRLRAENMLQLTENELH--RQER 351  
 DB 578 EKRR-RQEREROYRELEELRQEEQLNDRKLRQEEQLQREERELRQER 626

RESULT 15  
 TRHY\_SHEEP  
 ID TRHY\_SHEEP STANDARD: PRT: 1549 AA.  
 AC P22793;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-OCT-1994 (Rel. 30, last sequence update)  
 DT 15-JUN-2002 (Rel. 41, last annotation update)  
 DE Trichohyalin.  
 GN THH.  
 OS Ovis aries (Sheep).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecoridae;  
 OC Bovidae; Caprinae; Ovis.  
 OX NCBI\_TaxID=9940;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-93260018; PubMed-7684041;  
 RX Fietz M.J., McLaughlin C.J., Campbell M.T., Rogers G.E.;  
 RT Analysis of the sheep trichohyalin gene: potential structural and  
 RT calcium-binding roles of trichohyalin in the hair follicle.";  
 RT J. Cell Biol. 121:855-865(1993).  
 RN [2]  
 RP SEQUENCE OF 1016-1549 FROM N.A.  
 RC STRAIN-Merino-Dorset horn X Border Leicester; TISSUE-Wool follicles;  
 RX MEDLINE-90130632; PubMed-2298812;  
 RA Fietz M.J., Presland R.B., Rogers G.E.;  
 RT "The CDNA-decoded amino acid sequence for trichohyalin, a  
 RT differentiation marker in the hair follicle, contains a 23 amino acid  
 RT repeat.";  
 RL J. Cell Biol. 110:427-436(1990).  
 CC FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES  
 IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE  
 INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR  
 LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY  
 ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER  
 WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN

```

CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
CC DIFFERENTIATION.
CC -1- SUBUNIT: HOMODIMER (PROBABLE).
CC -1- ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS MAY BE PRODUCED BY
CC ALTERNATIVE SPLICING OF THE SAME GENE.
CC -1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
CC THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
CC THE EPITHELIA OF THE TONGUE, HOOF AND RUMEN.
CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED
CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
CC DIFFERENT SPECIES.
CC -1- PTM: SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE
CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE S-100
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: 218361; CAA79165.1; -
CC EMBL: X51695; CAA35992.1; -
CC PIR: A34209; A34209.
CC PIR: S32633; S32633.
CC PIR: A40691; A40691.
CC HSSP: P02633; 11G5.
CC InterPro: IPR001751; CaBP_S100.
CC InterPro: IPR002048; EF-hand.
CC Pfam: PF00036; efhand.1.
CC Pfam: PF01023; S_100.1.
CC ProDom: PD003407; CaBP_S100.1.
CC ProSite: PS00018; EF_HAND.1.
CC ProSite: PS00303; S100_CaBP.
CC Keratinization: Repeat: Calcium-binding.
CC
CC CA_BIND 22 33 EF-HAND 1 (LOW AFFINITY) (POTENTIAL).
CC CA_BIND 62 73 EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).
CC DOMAIN 413 832 14 X 28 AA APPROXIMATE TANDEM REPEATS.
CC REPEAT 413 448 1-1.
CC REPEAT 449 476 1-2.
CC REPEAT 477 504 1-3.
CC REPEAT 505 532 1-4.
CC REPEAT 533 560 1-5.
CC REPEAT 561 588 1-6.
CC REPEAT 589 616 1-7.
CC REPEAT 617 644 1-8.
CC REPEAT 645 678 1-9.
CC REPEAT 679 706 1-10.
CC REPEAT 707 742 1-11.
CC REPEAT 743 771 1-12.
CC REPEAT 772 796 1-13.
CC REPEAT 832 832 1-14.
CC DOMAIN 938 1507 23 X 23 AA APPROXIMATE TANDEM REPEATS.
CC REPEAT 938 961 2-1.
CC REPEAT 962 985 2-2.
CC REPEAT 986 1021 2-3.
CC REPEAT 1022 1044 2-4.
CC REPEAT 1045 1067 2-5.
CC REPEAT 1068 1090 2-6.
CC REPEAT 1091 1121 2-7.
CC REPEAT 1122 1144 2-8.

```

```

FT REPEAT 1145 1167 2-9.
FT REPEAT 1168 1197 2-10.
FT REPEAT 1198 1227 2-11.
FT REPEAT 1228 1250 2-12.
FT REPEAT 1251 1273 2-13.
FT REPEAT 1274 1296 2-14.
FT REPEAT 1297 1319 2-15.
FT REPEAT 1320 1342 2-16.
FT REPEAT 1343 1368 2-17.
FT REPEAT 1369 1391 2-18.
FT REPEAT 1392 1416 2-19.
FT REPEAT 1417 1439 2-20.
FT REPEAT 1440 1461 2-21.
FT REPEAT 1462 1484 2-22.
FT REPEAT 1485 1507 2-23.
FT VARIANT 1145 1197 MISSING (IN SHORT FORM).
FT VARIANT 1251 1273 MISSING (IN SHORT FORM).
FT CONFLICT 1399 1399 E -> G (IN REF. 2).
SQ SEQUENCE 1549 AA: 201173 MW: 87289 PFI326E54E CRC64:

```

```

Query Match 6.6% Score 125.5: DB 1: Length 1549:
Best Local Similarity 23.0% Pred. No. 5.2:
Matches 55: Conservative 42: Mismatches 71: Gaps 5:

```

```

QY 143 QQQRLGKKKKRRRPSKKRRHMKPYKLTWEKKKKFKDEKQSLRASRTIAEMFAKGPVAP 202
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 599 QEEQQLQROREKREKRRQRYLLEKVELDEEDYQROEKREKRRER----- 646
QY 203 YNTTFIMDDHDEPDLKTGLYSKRAAKSDOTSDDDFMERGGEGGSDMGDGSEF 262
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 647 ---QYLEKELOREERLO-----EEOQLREBERK-----RROE 677
QY 263 LQRFSEYERYHTESLQ-----NMSKQELIKEYLELEKCLSNMEDNNRLRLSKRLGSD 318
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 678 RERQYLEKVELDEEDYQROEKREKRRQRYLLEKVELDEEDYQROEKREKRRER----- 733
QY 319 DARVR-----ELELEDRLAENMLDTTEN-----ELHROGERAPLSK 356
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 734 DREKRRQVRERKYLELEDELOQEDRLOREK-QLLREDEKRRQYLEKVELDEEDYQROEK 791

```

Search completed: June 17, 2003, 10:22:45  
Job time : 28 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 10:21:00 ; Search time 26 Seconds  
(without alignments)  
406.263 Million cell updates/sec

Title: US-09-972-758A-2

Perfect score: 1910  
Sequence: 1 MAEPPLSEYHQHPOTSNGT.....LTENELHROGERAPISKFGD 359

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PCYUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	133	7.0	538	4 US-09-309-572-12	Sequence 12, Appl
2	133	7.0	1737	4 US-09-309-572-13	Sequence 13, Appl
3	131	6.9	1898	1 US-08-056-200-94	Sequence 94, Appl
4	131	6.9	1898	2 US-08-800-644-94	Sequence 94, Appl
5	125	6.5	754	4 US-09-214-564A-2	Sequence 2, Appl
6	123.5	6.5	557	4 US-08-979-608A-5	Sequence 5, Appl
7	123	6.4	1829	4 US-09-157-420-1	Sequence 1, Appl
8	122	6.4	723	1 US-07-814-964-11	Sequence 11, Appl
9	122	6.4	723	1 US-08-258-442-11	Sequence 11, Appl
10	122	6.4	723	1 US-08-328-809-6	Sequence 6, Appl
11	122	6.4	723	5 PCT-US92-11107-11	Sequence 11, Appl
12	121	6.3	538	4 US-08-370-368-9	Sequence 9, Appl
13	120	6.3	1780	1 US-08-769-309A-5	Sequence 5, Appl
14	120	6.3	1780	3 US-08-994-570-5	Sequence 5, Appl
15	118	6.2	530	4 US-09-562-737-39	Sequence 39, Appl
16	116	6.1	530	4 US-08-979-608A-8	Sequence 8, Appl
17	116	6.1	2972	4 US-09-579-181-1	Sequence 1, Appl
18	116	6.1	3118	4 US-09-579-181-1	Sequence 2, Appl
19	114.5	6.0	1958	1 US-07-945-283-2	Sequence 2, Appl
20	114	6.0	683	6 5210183-3	Patent No. 5210183
21	112.5	5.9	599	3 US-08-556-419-22	Sequence 22, Appl
22	112.5	5.9	629	3 US-08-556-419-23	Sequence 23, Appl
23	112.5	5.9	764	4 US-09-370-838-67	Sequence 67, Appl
24	112	5.8	802	4 US-09-156-316-1	Sequence 1, Appl
25	111	5.8	1162	2 US-08-728-323A-2	Sequence 2, Appl
26	111	5.8	1162	4 US-09-298-568-2	Sequence 2, Appl
27	110	5.8	576	2 US-08-533-306A-2	Sequence 2, Appl

28	110	5.8	576	2 US-08-742-923A-2	Sequence 2, Appl
29	108.5	5.7	657	3 US-08-893-852A-3	Sequence 3, Appl
30	108.5	5.7	657	4 US-08-821-818-3	Sequence 3, Appl
31	108	5.7	671	4 US-09-605-785-380	Sequence 380, App
32	108	5.7	671	4 US-09-439-313-380	Sequence 380, App
33	108	5.7	671	4 US-09-352-616A-380	Sequence 380, App
34	108	5.7	1115	2 US-08-568-459A-2	Sequence 2, Appl
35	108	5.7	1115	2 US-08-487-828B-2	Sequence 2, Appl
36	108	5.7	1115	4 US-09-210-288-2	Sequence 2, Appl
37	108	5.7	1115	6 5198347-6	Patent No. 5198347
38	108	5.7	1719	4 US-09-605-785-378	Sequence 378, App
39	108	5.7	1719	4 US-09-439-313-378	Sequence 378, App
40	108	5.7	1719	4 US-09-352-616A-378	Sequence 378, App
41	108	5.7	2509	2 US-08-149-097D-35	Sequence 35, Appl
42	107.5	5.6	1164	4 US-09-457-708-2	Sequence 2, Appl
43	107.5	5.6	1637	4 US-09-718-692-2	Sequence 2, Appl
44	107.5	5.6	1637	4 US-09-718-852-2	Sequence 2, Appl
45	107.5	5.6	1637	4 US-09-718-815-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-09-309-572-12  
: Sequence 12, Application US/09309572  
: Patent No. 6440730  
: GENERAL INFORMATION:  
: APPLICANT: Heinrich-Pette-Institut  
: TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV  
: FILE REFERENCE: P50489  
: CURRENT APPLICATION NUMBER: US/09/309,572  
: EARLIER FILING DATE: 1995-05-11  
: EARLIER APPLICATION NUMBER: DE 198 56 463  
: NUMBER OF SEQ ID NOS: 24  
: SOFTWARE: PatentIn Ver. 2.0  
: SEQ ID NO 12  
: LENGTH: 538  
: TYPE: PRT  
: ORGANISM: Moloney murine leukemia virus  
: FEATURE:  
: OTHER INFORMATION: gag protein  
US-09-309-572-12

Query Match	7.0%	Score 133:	DB 4:	Length 538:
Best Local Similarity	20.9%	Pred. No. 0.0019:		
Matches	89:	Conservative	40:	Mismatches 142: Indels 154: Gaps 15:
QY	30	PERPPGA-----EERVPEDSRMOSRARPOLGRRGPE---	CEGS-----LESQPPPL	74
DB	107	PPLPPSPSLPLEPPRSTPPRSSLLYPALE-PSLGAKKPPVLSDSGPIIDLITDDPPV		165
QY	75	QTACPESSCLREGCKGQNDSSAGGDPPEPAEVEFPFAELL---	AQCHDSASKL	130
DB	166	RDRPPPSD---RDG---NGGEATPAGEAP-----DPSPASLRRRREPPADSTTTSA		214
QY	131	GAPAAEGEEMGOQOOLGKKHRRRPSKKRWKPYLLTWEEKKPPKOSLRASRIR		190
DB	215	FPLRAGN-----COLQYWPSSSDLYNMK-----NNNPSFSDPGKLTALTE		257
QY	191	AEMFAGQPVAPYNTTQFLIMDDHDQEEPDLKTQ-----LYSKRAAKGSD-----		235
DB	258	SVLL-----THQPTWDDOQOQLGLTLTLTGECKORVLLERAKAVRDDDRPTQLPNE		307
QY	236	-----TSDDFMEEGGE-----	DGSDMGDGGE	261
DB	308	VAAAPLEPPDDYTTQAGRNHLVHYROLLLAGLONAGRSPTNLAKVKGITGPPNPSPA		367
QY	262	FLORDSEFTYERY-----HTESLONNSKDELLKE		290
DB	368	FLER-LKEATYRRTTPDPEDPGQETNVSMSTWOSAPDIGRKLRLDLKNTIGDLVR-		425

```

OY      291 YLETCLECEENNRRLTESKRLGGDDAR-RELELEIDRLAENJOLLTENEHLRQ 349
Db      426 --EAEKIFPKRRETPER-----ERIRRETEKEERRRTEDEQKEKENDRRRR 472

OY      350 ERAPL 354
Db      473 EMSKL 477

RESULT 2
US-09-309-572-13
: Sequence 13, Application US/09309572
: Patent No. 6440730
:
: GENERAL INFORMATION:
: APPLICANT: Heinrich-Pette-Institut
: TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV
: FILE REFERENCE: P50489
: CURRENT APPLICATION NUMBER: US/09/309,572
: EARLIER FILING DATE: 1999-05-11
: EARLIER APPLICATION NUMBER: DE 198 56 463
: NUMBER OF SEQ ID NOS: 24
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 13
:
: LENGTH: 1737
: TYPE: PRT
: ORGANISM: Moloney murine leukemia virus
: FEATURE:
: OTHER INFORMATION: gag-pol protein
US-09-309-572-13

```

```

Query Match Similarity      7.0%: Score 133; DB 4; Length 1737;
Best Local Similarity       20.9%: Pred No. 0.0088;
Matches      89: Conservative    40: Mismatches   142: Indels    154: Gaps     19

```

	QY	DB	107	75	166	QY	131	215	QY	Db	236	QY	308	262	368	QY	291	426	QY	350	Db	473
	PERPGA-----EERVPEEDSRMOSRAFPOLGRPGDE--GECS-----LESOPPL 74		PPLPSASLPLEPRSTRPPRSSLYPALT-PSLGAKPKQVUSDGGPILDLITBDDPPY 165	QTQAPESSCLREGEKGNGDSDSAGGDPPEPAVEVPPEAEELL---AQCNDSEASKL 130	KDPRPPSPD--RDG----NGCAATAGAP-----DSPSMASRLGRRPREPYADSTTSGA 214		GAPAAGEEWGQQOQROLKKKHRRRPSKKNHMKPYKLTWEKKKFDEKOSLASAIR 190	FLRAGGN-----GQLQWPFSSSDLDYNWK-----NNNPFSEDEPKULTALIE 257					VDAEPLEPRDMDYTTQAGRNLHVHYROLLLAGLONAGRSPYNLAVKKITGPNESPFA 367	PIORPFSERYKY-----HTSESJONSKOELIKE 290	FIER-LKEAYRRRTYPDPEDPGQETNVMSFTIWSQAPDLGRKLERIEDLKNTIGLVLR- 425		YLELEKLSRMEDENNRLRLESKRILGGDARY-RELELEDRLRAENLOLLTENELHRQO 349	--EAEEKIFMKRETPEER-----EBIRIRETEEKEEKERRRTDEDOKEREKDRRRRH 472		ERAPL 354		EMSKL 477

```

RESULT 3
US-08-056-200-94
; Sequence 94, Application US/08056200
; Patent No. 5616500
; GENERAL INFORMATION:

```

APPLICANT: Steinert, Peter M.  
APPLICANT: Lee, Seung-Chul  
APPLICANT: Kim, In-Gyu  
APPLICANT: Chung, Soo-Il  
APPLICANT: Park, Sang-Chul  
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
TITLE OF INVENTION: Methods of Using Same  
NUMBER OF SEQUENCES: 117  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/056, 200  
FILING DATE: 30-APR-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Fedrick, Michael F.  
REGISTRATION NUMBER: 36,799  
REFERENCE/DOCKET NUMBER: NH054.001A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (714) 760-0404  
TELEFAX: (714) 760-9502  
INFORMATION FOR SEQ ID NO: 94:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1898 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-056-200-94

```

Query Match          6.9%  Score 131:  DB 1:  Length 1898:
Best Local Similarity 22.4%:  Pred. No. 0.015:
Matches 79:  Conservative 54:  Mismatches 146:  Indels 74:  Gaps 13:

Oy      31  ERPPGAER-----VPEEDSRMOSRAFPOLGSRPGPEGSGSLSEOPPLQTAGCESS 83
      ||| |||
      ||| |||
      ||| |||
Db      813  ERLRAERQREQREFLPLEEKEQR-----GROREREKEQLFLEEQLORERRAQ 865

Oy      84  CLREGEKQNDSDSAGDFPPPAVEVTPAE-----LLAOPCHSESKLGAAPAQ 136
      ||| |||
      ||| |||
      ||| |||
Db      866  QLOEEDLDLOEQRERROEQRRDQWRMQLTEERKRRRHTLYAKPALOEQLRKEOQLLE 925

Oy      137  GEEWEGOOQOLGKKKHHRRRSKKRHKKPYUKLTWEKKKPFDEKOSLRASIRAEWFAK 196
      ||| |||
      ||| |||
      ||| |||
Db      926  EEEELQREERE-----KKRQEOEPROYREEEQLQOEEOILLREERKKRRQERERYRK 979

Oy      197  GQVAPRYNTQFLMDHDQ--EEFDLTGLYSKRAAKSDPDSDDPFMEGGEEDGSD 253
      ||| |||
      ||| |||
      ||| |||
Db      980  DKK-----LOOKEBQLLGEPEEK--RROEREKKYREEEBLQOE----- 1016

Oy      254  GMGDSGSEFLQDFEYTER-----YHTESLONNSKQELIKYELEKLSRMD--EN 305
      ||| |||
      ||| |||
      ||| |||
Db      1017 -----EEOILLREERKKRRROEWEROYKKKDLQOEEOILLREERKKRLOEREQYREE 1070

Oy      306  NLRLESRLGDDARVR---ELELEL---DLRAENQLITEN-ELARQOER 351
      ||| |||
      ||| |||
      ||| |||
Db      1071  EELQOEEOILLGEERETRRQELLENQYKKLEELQOEEOILLREERKKRRQER 1123

RESULT 4
US-08-800-644-94
; Sequence 94, Application US/08800644
; Patent No. 5958752
; GENERAL INFORMATION:

```

```

: APPLICANT: Steinert, Peter M.
: APPLICANT: Lee, Seung-Chul
: APPLICANT: Kim, In-Gyu
: APPLICANT: Chung, Soo-Il
: APPLICANT: Park, Sang-Chul
: TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
: TITLE OF INVENTION: Methods of Using Same
: NUMBER OF SEQUENCES: 117
: CORRESPONDENCE ADDRESS:
: ADDRESS: Knobbe, Martens, Olson & Bear
: STREET: 620 Newport Center Drive, Sixteenth Floor
: CITY: Newport Beach
: STATE: CA
: COUNTRY: U.S.A.
: ZIP: 92660
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/800,644
: FILING DATE: 14-FEB-1997
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/056,200
: FILING DATE: 30-APR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Fedtich, Michael F.
: REGISTRATION NUMBER: 36,799
: REFERENCE/DOCKET NUMBER: NIH054.001A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (714) 760-0404
: TELEFAX: (714) 760-9502
: INFORMATION FOR SEQ ID NO: 94:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1898 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-800-644-94

```

```

Query Match      6.9%; Score 131; DB 2; Length 1898;
Best Local Similarity 22.4%; Pred. No. 0.015;
Matches 79; Conservative 54; Mismatches 146; Indels 74; Gaps 13;

```

```

QY 31 ERPPGAEER-----VPEEDSRMOSRAFPOLGRRPGEBSLESOPPLQTOACPBSS 83
DB 813 EROLRAEEROOREGRLPEEEKEOR-----GRORREKELQFLEEFOLQRRERAQ 865
QY 84 CLRGEGGONGDDSSAGDFPPPAVEPTPEAE-----LLAOPCHDSEASKLGADPAAG 865
DB 866 QLOEEEGLOEDORRROEOORROOKWMOLEEFKKRRHNTLYAKPALQOLRKQOOLQE 925
QY 137 GEEFWGQOQOLGKKHRRRPSKKRHWKYKLTWEEKKKFKDEKOSTIRASRIAEWFAK 196
DB 926 EEEELQREERE-----KRRROEOEROYREEOLOQEEFOLLREERERRRROEROYRK 979
QY 197 GQAPAPYNTTOFLMDHDQ---FEEDLKTGLYSKRAAKSDTSDDDFMEGGEGDGGSD 253
DB 980 DKR-----LQKKEQLLGEPEERK-----RROERKKYREEELOQE----- 1016
QY 254 GMGDGSSEFLQDRDSEYER-----YHTESLQMSQOELIKYLELEKCLSRMED--EN 305
DB 1017 -----EQLLRERERKRROEMEROYRKDELQOEEQOLLREERKRRLDEREROYREE 1070
QY 306 NRLEFKRLGGDDARV---ELELEL---DLRAENMLQLTEN-ELARQOER 351
DB 1071 EELQOEEELIGERETRRROELEROYRKDEELOQEEFOLLREPEERKRROER 1123

```

RESULT 5  
US-09-214-564A-2

```

: Sequence 2, Application US/09214564A
: Patent No. 6150515
: GENERAL INFORMATION:
: APPLICANT: Sharp, Phillip A.
: APPLICANT: Zhou, Qiang
: TITLE OF INVENTION: TAT-SF: Cofactor For Stimulation Of Transcriptional
: TITLE OF INVENTION: Elongation By HIV-1 TAT
: FILE REFERENCE: M0656/7042
: CURRENT APPLICATION NUMBER: US/09/214,564A
: PRIOR APPLICATION NUMBER: US 60/021,218
: PRIOR FILING DATE: 1996-07-03
: PRIOR APPLICATION NUMBER: US 60/033,152
: PRIOR FILING DATE: 1996-12-13
: PRIOR APPLICATION NUMBER: PCT/US97/11713
: PRIOR FILING DATE: 1997-07-03
: NUMBER OF SEQ ID NOS: 5
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 2
: LENGTH: 754
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-214-564A-2

```

```

Query Match      6.5%; Score 125; DB 4; Length 754;
Best Local Similarity 21.5%; Pred. No. 0.014;
Matches 82; Conservative 50; Mismatches 166; Indels 84; Gaps 17;

```

```

QY 9 YOHOPOTSNCGAAYOEELNPPRPGAEEVPEEDSRMOSRAFPOLG--RGPGEBS 66
DB 401 FSEHPSTSKMNA-----QETATGMARFEEPIDEKKFKP---TEDGGEFEESGSENN 448
QY 67 LESOPPLQTOACQ-----ESSCLREGEKGGONGDDSSAGPPPAVEPTPEAE---LL 118
DB 449 KESSPEKEAEBGCEKSEBGRGEGSCQKESSEGN-PRGSEDSPPKESKKKTL 507
QY 119 AOPCHDSEASKLGADPAAGGEEWGOQOLGKKHRRRPSKKRHWKYKLTWEEKKF 178
DB 508 KNDCEENGKLAKESDDLINKSE-----EVGPIKESSEDSSEK-----ESPDG 551
QY 179 DEKOSLRASRIAEWFAKGPVAPYNTTOFLMDHDQ---EPDLKTGLYSKRAAKSDT 236
DB 552 SEKQSEDSEREFEF-----ENGLKXLDDEGSEKELHENVLKE--LEENDS 556
QY 237 SDDFMEEGGEFNOGSGMGGDGSSEFLQDRFSE-----TYER-YHTESLQNSKQ 285
DB 597 ENSEF-----EDDSSEKVLDEEGSE---REFDESDKEEEDDYEXKVFDESDKEDEE 648
QY 286 ELIKYLELEKCLSRMEDENNRLRLESKRRLGGDARVRELELELRLRAE---NLQLTTE 342
DB 649 YADKGLGLADKKRAEBBDADKLFEEESDQKDEDEDADGKEVADADEKLFEDDDSNKLFDE 708
QY 343 NE-----LHROERAPLSKFG 358
DB 709 EEDSEKRLFPDSDRGTLGGFG 730

```

```

RESULT 6
US-08-979-608A-5
: Sequence 5, Application US/08979608A
: Patent No. 6355451
: GENERAL INFORMATION:
: APPLICANT: Lees, Ann M.
: APPLICANT: Lees, Robert S.
: APPLICANT: Law, Simon W.
: APPLICANT: Arjona, Anibal A.
: TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
: BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
: TREATING ATHEROSCLEROSIS
: NUMBER OF SEQUENCES: 42
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 225 Franklin Street

```

CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02110-2804  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/979,608A  
 FILING DATE: 26-NOV-1997  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/048,547  
 FILING DATE: 03-JUN-1997  
 APPLICATION NUMBER: US 60/031,930  
 FILING DATE: 27-NOV-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Myers, Louis  
 REGISTRATION NUMBER: 35,965  
 REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617/542-5070  
 TELEFAX: 617/542-8906  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 557 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 FRAGMENT TYPE: internal  
 SEQUENCE DESCRIPTION: SEQ ID NO: 5:

US-08-979-608A-5

Query Match 6.5%: Score 123.5; DB 4; Length 557;  
 Best Local Similarity 22.6%: Pred. No. 0.013;  
 Matches 91; Conservative 53; Mismatches 123; Indels 135; Gaps 20;

20 GAAVQELNPPGAEERVP-----EDSRWQRAFPOLGRRPGEGLSEQ 70  
 9 GAA--KQPNKSSPGQPEAGAGQGRPPAPAREAGASSQAPGR--PEGAAQAKYQ 63  
 71 PPPL-----QTQACPESSCLREGKQNGDDSSAGDPFPAEVEPTPEALLAOP 121  
 64 PGLADVSEELSRQLELDLITYCV--DNNQAGPGEDEVG-----EP--PEPE---- 107  
 122 CHDSASKL-----GAPAAAGEEEMGOQ-----POLGKKRRRRRPSKKR 162  
 108 --DAKSRAYVARNGEPEPTPVVNGEKETSKAEGTEIITSDEVGDRDHRROEKKA 165  
 163 HW-----KPYKLTWEKKKFKDEKQSLRASIRAFKAGQPVAPYNTT 206  
 166 KGLCKEITLMTNTLSTPEKIALCKKYAEELLEHRNSQKMKLLQKKQ-----S 218  
 207 QFLDDHDQEPDRLKTLGYSKRAAKSDDTS-----DDDFEEGGEEDGSGDMGCD 258  
 219 QLV-----QEKDLRGHSHKAILARSKLESLELQRLHNSLKEEG----- 259  
 259 GSEFLQRFSELYRYHTESLQNSKQELIKLEYLEKLSRMEDENNRLRLESKRIGCD 318  
 260 ---VQRRAREEERKKETVSHFQMTLNDI---QLQME---QHNERNSKLLQENMELA-- 306  
 319 DARVRELELDRLRAENL-----QLTENEHLHROE 350  
 307 -ERLKL-IEOYELREHIDKVFKKHKLQOOLVDAKLQAOE 346

RESULT 7  
 US-09-157-420-1  
 Sequence 1, Application US/09157420  
 Patent No. 6180760  
 GENERAL INFORMATION:  
 APPLICANT: TAKAI, Yoshimi

APPLICANT: NAKANISHI, Hiroyuki  
 APPLICANT: MANDAI, Kenji  
 APPLICANT: WADA, Manabu  
 APPLICANT: OBAISHI, Hiroshi  
 TITLE OF INVENTION: ACTIN FILAMENT-BINDING PROTEIN "L-AFADIN"  
 FILE REFERENCE: 98-1042\*/LC(WMC)/653  
 CURRENT APPLICATION NUMBER: US/09/157,420  
 CURRENT FILING DATE: 1998-09-21  
 NUMBER OF SEQ ID NOS: 1  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO: 1  
 LENGTH: 1829  
 TYPE: PRT  
 ORGANISM: rat  
 US-09-157-420-1

Query Match 6.4%: Score 123; DB 4; Length 1829;  
 Best Local Similarity 19.9%: Pred. No. 0.068;  
 Matches 83; Conservative 59; Mismatches 157; Indels 118; Gaps 15;

9 YOHQOTSNCGAAV-----QELNPPGAEERVP-----EDSRWQRAFP 53  
 1259 YEEKPHMTESDHSIAIORYTRSOELREKVVOLERHVESGMDKCDSDMTNQS-- 1316  
 54 QLGRRPGEGLSESOPEPIQTOACPESSCLREGKQNGDDSSAGDPFPAEVEPTP 113  
 1317 ---SSVESSSTSSQEHNLHNSKSVTPASTLTKSP-----GRMKTAAVLPPT 1360  
 114 EAELLAPCH-----DSEASKIGADPAAGEEEMGOQROLGKKHRR 155  
 1361 VA--VSQPIRTDLPPPPPPAHYTSDFGISMDELPLPPPPANA-AQSAQVAAAEKK 1417  
 156 PPSKKRRHWPYKLTWE-EKKKFKDEKQSLRASIRAFKAGQV----- 201  
 1418 REEHQRWYEEKARLEEREKRRQEKLLQMTQSLNPASFPLAQAPKRPSTLQ 1477  
 202 PYNTQFLMDHDEP-----DLKTGLYSKRAAKSDDTSDDDFMEGGEEDGSGDM 255  
 1478 PQETV--IRELQPOQPPRTERRDLQYTTISKEELSDGSLSPBW----- 1521  
 256 GGDSEFLQRFSELYRYHTESLQNSKQELIKLEYLEKLSRMEDENNRLR----- 309  
 1522 ---KRDAREKLEKQQGMHIVDM---LSKEIHQLQKKGRTAESDRLKRLMEW 1569  
 310 ---LESKRLGCDDAVRELELE---LDRLAENLQLTENEHLHROERAPLSK 356  
 1570 QFQKRLQESKKKDDDEDEEDDDVDYTLIMQRLAENRRARLQDEERRQQLLEKMK 1626

RESULT 8  
 US-07-814-964-11  
 Sequence 11, Application US/07814964  
 Patent No. 5359047  
 GENERAL INFORMATION:  
 APPLICANT: Donahue, Brian A.  
 APPLICANT: Toney, Jeffrey H.  
 APPLICANT: Bruhn, Suzanne L.  
 APPLICANT: Pili, Pieter M.  
 APPLICANT: Brown, Steven  
 APPLICANT: Kellett, Patti  
 APPLICANT: Essigmann, John M.  
 APPLICANT: Lippard, Stephen J.  
 TITLE OF INVENTION: DNA Structure Specific Recognition  
 NUMBER OF SEQUENCES: 13  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
 STREET: 2 Militia Drive  
 CITY: Lexington  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02173  
 COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/814,964
FILING DATE: 19911226
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/539,906
FILING DATE: 18-JUN-1990
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-4787AAA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 11
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 723 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Drosophila melanogaster
IMMEDIATE SOURCE:
CLONE: Drosophila SSRP (predicted)
FEATURE:
NAME/KEY: Domain
LOCATION: 458..507
OTHER INFORMATION: /label= Acidic
FEATURE:
NAME/KEY: Domain
LOCATION: 518..547
OTHER INFORMATION: /label= Basic I
FEATURE:
NAME/KEY: Domain
LOCATION: 547..620
OTHER INFORMATION: /label= HMG-box
FEATURE:
NAME/KEY: Domain
LOCATION: 632..649
OTHER INFORMATION: /label= Basic II
FEATURE:
NAME/KEY: Domain
LOCATION: 657..723
OTHER INFORMATION: /label= Mixed Charge
US-07-814-964-11

```

```

Query Match          6.4%; Score 122; DB 1; Length 723;
Best Local Similarity 20.4%; Pred. No. 0.025; Indels 58; Gaps 9;
Matches 61; Conservative 55; Mismatches 125;

```

```

QY 63 GGSLSQPPPLQTOACPESSCLREGKQNGDDSSAGGDFPP---PAVEPTPEAEIIA 119
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 441 GDSNDNEDDAYLARLKAFAREKEDDDDDGSDSESTDDFKPNMESVVAEYDSNVSS 500
QY 120 QPCHSSEASKLAPAAGEEENGQOQOLGKKKRRRRPSKKRHKMPYKLTWEKKKFD 179
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 501 DSDDDSDAS---GGGSDSGAKKKE--KKSEKKKKKKKH-KEKERTKPPSKKKD 551
QY 180 EKOSLRASIRAEAFKGPVAPYNTQFLMDHD-----QEEPLKTGLISKR----- 228
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 552 SGRPKRA-----TTATMLWLNDTRISIKRNPQIKVTETLAKKGEMW 593
QY 229 -----AAASDDTSDDDFMEEGEGEDGSDMGDSEFLQRFSETYEYRHT 277
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 594 KELKDKSKWDAAKKKQRYHDEMRYKFEAGGDSNKGSGSKKRKTEPSSKKANTS 653
QY 278 SLQNNSKOFLIKEYLELEKCLSRMEDENNR-LRLESKRUGGDARVRELELELRLRAE 335
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

```

```

Db 654 GSGFKS-----KEYISDDSTSSDDEKONEPAKKSKPPSDGAKKKAKKSESEPESE 707
RESULT 9
US-08-258-442-11
: Sequence 11, Application US/08258442
: Patent No. 5670621
: GENERAL INFORMATION:
: APPLICANT: Donahue, Brian A.
: APPLICANT: Toney, Jeffrey H.
: APPLICANT: Bruhn, Suzanne L.
: APPLICANT: Pil, Pieter M.
: APPLICANT: Brown, Steven
: APPLICANT: Kelleet, Patli
: APPLICANT: Essigmann, John M.
: APPLICANT: Lippard, Stephen J.
: TITLE OF INVENTION: DNA Structure Specific Recognition
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
: STREET: 2 Militia Drive
: CITY: Lexington
: STATE: MA
: COUNTRY: USA
: ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/258,442
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/539,906
FILING DATE: 18-JUN-1990
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-4787AAA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 11:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 723 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Drosophila melanogaster
IMMEDIATE SOURCE:
CLONE: Drosophila SSRP (predicted)
FEATURE:
NAME/KEY: Domain
LOCATION: 458..507
OTHER INFORMATION: /label= Acidic
FEATURE:
NAME/KEY: Domain
LOCATION: 518..547
OTHER INFORMATION: /label= Basic I
FEATURE:
NAME/KEY: Domain
LOCATION: 547..620
OTHER INFORMATION: /label= HMG-box
FEATURE:
NAME/KEY: Domain
LOCATION: 632..649
OTHER INFORMATION: /label= Basic II
FEATURE:

```









```
Db 307 SESEPEPELHETPRRPAFLPVVYDDTNSYEWGSESEPDLSYDADSPWLLSA--LVSr 363
QY 59 PCPEGE-----GSLESOPPLQTOACPESSCLREGKCGONGDSSAGDF----- 103
Db 364 MISEGDSPIKCPQOCSPAPRL-----PEFFASQANSVPQOCODPEAGPHHELVDMDTLC 418
QY 104 --PPPA-----EVEPTPEALLAQPCHDSEASKLGAPAAAGSEEWGQOQOLGKKKRR 154
Db 419 IPPPPAAPAAKLGPAQGP-CMFLSNPTRDNTNPLMATP-----GRQARP-GRCSA 468
QY 155 RRPSSKKRRHKPYKLTWEKKKFKDEKQSLRA-----SRIAEHFAKGQPYAP---- 202
Db 469 RCSEEE-----EEDESEDEDEEDATDSVVPGSRVYTG-----TAPLDaw 509
QY 203 --YNTTQ--FLMDHDOEEPDLTGLYSKRAAKSDOTSDDDFMEEGGEEDGSDMGCD 258
Db 510 LVYDAVKYTYVDEHTQLE-----AVSLRRCAGLNDSEEDSSCFASEEAGATGLGSD 563
QY 259 G-----SEFLQRPDSERYRYHTESLQNM-----KOELIKEYLELEKCLS 299
Db 564 QVEDHSPDSPDLTFIKKFLNVFVNKTSRSSSTESLGLFSCVYXNGMERBQTHRAVNRFTP 623
QY 300 RMEDENNRLRLKESKRUGDDARYBELE 326
Db 624 RHPDE-----QELDVDDPVLLREAE 642
```

Search completed: June 17, 2003, 10:25:34  
Job time : 29 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comphen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: June 17, 2003, 10:55:42 ; Search time 194 Seconds  
(without alignments)  
2679.659 Million cell updates/sec

Title: US-09-972-758A-2  
Perfect score: 1910  
Sequence: 1 MAEPFLSEYOHQPTSNCTG.....LTENELHROQERAPLSKFGD 359

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 1029858 seqs, 724030393 residues

Total number of hits satisfying chosen parameters: 2059716

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n model -DEV=xlh  
-Q=/cgn2.1/USPTO\_spool/US09972755/runat\_17062003\_095239\_24438/app.query.fasta\_1.519  
-DB=publshed\_applications\_NA -QFMT=fastacp -SUFFIX=rrpb -MIMATCH=0.1  
-LOOPEXT=0 -LOOPEXT=0 -ONITS=bits -START=1 -END=1 -MATRIX=blousum62  
-TRANS-human40.cdi -LIST=45 -LOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALICN=15 -MODELOC=1 -OUTFMT=ptc -NOR=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USPR=US09972758.@CGN.1.1.57.@runat\_17062003\_095239\_24438  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONGILOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: /cgn2.6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2.6/ptodata/1/pubpna/PCT\_NEM\_PUB.seq:\*  
3: /cgn2.6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2.6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2.6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2.6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2.6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2.6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2.6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2.6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2.6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2.6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2.6/ptodata/1/pubpna/US60\_NEM\_PUB.seq:\*  
14: /cgn2.6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1910	100.0	2199	10	US-09-745-763-33
2	1902	99.6	1080	9	US-09-972-758-1
3	691	36.2	414	10	US-09-983-965-4895
4	629	32.9	461	9	US-09-954-531-813

Result No.	Score	Query Match	Length	DB ID	Description
5	619	32.4	495	9	US-09-918-995-403
6	184	9.6	549	9	US-09-991-936-1712
7	152	8.0	2421	10	US-09-974-300-7891
8	147.5	7.7	3773	10	US-09-925-302-47
9	147	7.7	7453	9	US-10-037-270-248
10	147	7.7	7501	9	US-10-037-270-249
11	145	7.6	4301	10	US-09-917-800A-1666
12	141	7.4	2962	9	US-09-291-417-106
13	139	7.3	6537	9	US-10-175-522-186
14	138.5	7.3	5835	9	US-09-927-597-1
15	138.5	7.3	5937	9	US-09-927-597-3
16	138.5	7.3	6861	9	US-10-171-311-161
17	138.5	7.3	6900	9	US-10-171-311-163
18	138	7.2	1960	10	US-09-834-975-833
19	137	7.2	8332	10	US-09-006-298-1
20	135.5	7.1	2654	9	US-10-050-704-40
21	135.5	7.1	3476	9	US-10-245-103-51
22	135.5	7.1	3476	9	US-10-245-107-51
23	135.5	7.1	3476	9	US-10-245-143-51
24	135.5	7.1	3476	9	US-10-245-171-51
25	135.5	7.1	3476	9	US-10-245-851-51
26	135.5	7.1	3476	9	US-10-245-883-51
27	135.5	7.1	3476	9	US-10-237-535-51
28	135.5	7.1	3476	9	US-10-238-183-51
29	135.5	7.1	3476	9	US-10-238-283-51
30	135.5	7.1	3476	9	US-10-238-370-51
31	135.5	7.1	3476	9	US-10-245-055-51
32	135.5	7.1	3476	9	US-10-245-147-51
33	135.5	7.1	3476	9	US-10-245-730-51
34	135.5	7.1	3476	9	US-10-245-739-51
35	135.5	7.1	3476	9	US-10-246-210-51
36	135.5	7.1	3476	9	US-10-239-196-51
37	135.5	7.1	3476	9	US-10-243-024-51
38	135.5	7.1	3476	9	US-10-243-409-51
39	135.5	7.1	3476	9	US-10-245-033-51
40	135.5	7.1	3476	9	US-10-245-621-51
41	135.5	7.1	3476	9	US-10-245-880-51
42	135.5	7.1	3476	9	US-10-243-095-51
43	135.5	7.1	3476	9	US-10-243-185-51
44	135.5	7.1	3476	9	US-10-245-427-51
45	135.5	7.1	3476	9	US-10-245-473-51

#### ALIGNMENTS

RESULT 1  
US-09-745-763-33  
; Sequence 33, Application US/09745763  
; Patent No. US20020065394A1  
; GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth

McCoy, John M.  
Lavallee, Edward R.  
Collins-Racie, Lisa A.  
Evans, Cheryl  
Merberg, David  
Trecay, Maurice  
Spaulding, Vikki

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
ENCODING THEM

NUMBER OF SEQUENCES: 219

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: MA

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

```

:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/745,763
: FILING DATE: 18-Jun-2000
: CLASSIFICATION: <Unknown>
: ATTORNEY/AGENT INFORMATION:
:   NAME: Sprunger, Suzanne A.
:   REGISTRATION NUMBER: 41,323
: TELECOMMUNICATION INFORMATION:
:   TELEPHONE: (617) 498-8284
:   TELEFAX: (617) 876-5851
: INFORMATION FOR SEQ ID NO: 33:
:   SEQUENCE CHARACTERISTICS:
:     LENGTH: 2199 base pairs
:     TYPE: nucleic acid
:     STRANDEDNESS: double
:     TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: SEQUENCE DESCRIPTION: SEQ ID NO: <Unknown>
:
:
: US-09-745-763-33
:
:
: Alignment Scores:
: Pred. No.: 1 22e-167 Length: 2199
: Score: 1910.00 Matches: 359
: Percent Similarity: 100.00% Conservative: 0
: Best Local Similarity: 100.00% Mismatches: 0
: Query Match: 100.00% Indels: 0
: DB: 10 Gaps: 0
:
: US-09-972-758a-2 (1-359) x US-09-745-763-33 (1-2199)
:
: QY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
:   |||||||
: Db 707 ATGGCCGAGCCATCTTCTGTCAAGATATCAACACACACCTCAAACTGACATCTACAGGT 766
:
: QY 21 AlaAlaAlaValGlnGlnGlnLeuAsnProGluIleuArgProProGlyAlaGluGluVal 40
:   |||||||
: Db 767 GCGTGGCTGTCCAGGAAGAGCTGAACCTGAGCGCCCGCCAGGCGGAGAGCGGGTGG 826
:
: QY 41 ProGluGluAspSerArgTTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
:   |||||||
: Db 827 CCGAGAGAGACAGTGTGGTGGCATGTGAAGCGCTTCCCGCAGTGGTGGTGGTGGTGG 886
:
: QY 61 ProGluGluGlyGlySerLeuGlnSerGlnProProProLeuGlnThrGlnAlaCysPro 80
:   |||||||
: Db 887 CCGAGAGGAGGAGGAGGAGGAGTCCCAACCAACCTCCCTGCAGACCCAGCGCTGTCCA 946
:
: QY 81 GluSerSerCysLeuArgGlnGlyGlyGlyGlnAsnGlyAspAspSerSerAlaGly 100
:   |||||||
: Db 947 GAATCTAGCTGCTGTGAGAGAGGCGGAGAAAGGCGCAATGGGAGCGACTGTCTGCTGGC 1006
:
: QY 101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
:   |||||||
: Db 1007 GCGCACTTCCCGCGCGCGCGAGAGTGAACCGAGCCCGGAGCGGAGCTGCTCGCCGCG 1066
:
: QY 121 ProCysHisAspSerGlnAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGlnGlu 140
:   |||||||
: Db 1067 CCTTGTCATGACTCCGAGGCGCAGTAAAGTGGGGGCTCCCGCGCAGGCGGCGCAAGAGAG 1126
:
: QY 141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
:   |||||||
: Db 1127 TGGGGACAGACAGACAGACAGCTGGGGAGAAAGAAACATAGGAGAGCGCCGCTCCAGAG 1186
:
: QY 161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
:   |||||||
: Db 1187 AAGCGGCATTTGGAACCCGTAACAAGCTGACCTGGGAAGACAGAAAGTTCCGACGAG 1246
:
: QY 181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
:   |||||||
: Db 1247 AAACAGAGGCTTCGAGCTTCAAGAGATCCGAGCCGAGATGTTCGCCAAGGCGCAGCGGTC 1306
:
: QY 201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGlnGluProAspLeu 220
:   |||||||
: Db 1307 GGGCCCTATTAACACACGAGCTTCTCATGATGATACGACCGAGAGAGCGCGGATCTTC 1366
```

```

:
: QY 221 LysThrGlyLeuTyrSerLysArgAlaAlaLysSerAspPThrSerAspAsp 240
:   |||||||
: Db 1367 AAACCGGCGCTTACTTCCAGAGCGGGCGCCGCCCAATTCGACAGACACGCGATACGAC 1426
:
: QY 241 PheMetGluGlnGlyGlyGlnLeuAspGlyGlySerAspGlyMetGlyLysArgLys 260
:   |||||||
: Db 1427 TTCATGGAAGAGAGGGGTGAGAGGATGGGGGACGAGATGGATGGAGAGGAGCGGAC 1486
:
: QY 261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGlnSerLeuGln 280
:   |||||||
: Db 1487 GAGTTTCTCCACCGGACTTCTCGGAGAGTACGAGCGGTACACAGGAGCGCTGCAG 1546
:
: QY 281 AsnMetSerLysGlnGlnLeuLeuLysGlyTyrLeuGlnLeuGlnLysCysLeuSerArg 300
:   |||||||
: Db 1547 AACATGACGACAGACAGAGCTCATCAAGAGACTACCTGGAACCTGAGAAAGTCTTCGCC 1606
:
: QY 301 MetGluAspGluAsnAsnArgLeuArgLeuGlnSerLysArgLeuGlyLysAspAla 320
:   |||||||
: Db 1607 ATGGAGAGACGAGAACACACCGGCTGGCGGCTGGAGAGCAGCGGCTGGTGGCAGCGG 1666
:
: QY 321 ArgValArgGlnGlnGlnGlnLeuGlnLeuAspArgLeuArgAlaGluAsnLeuGlnLeu 340
:   |||||||
: Db 1667 CGTGTGGGAGCTGTGAGCTGTGACCGGCTGCGCGCGGAGAACCTCCAGCTGCTG 1726
:
: QY 341 ThrGluAsnGlnLeuHisArgGlnGlnGlnIleuArgAlaProLeuSerLysPheGlyAsp 359
:   |||||||
: Db 1727 ACCGAGACGAACTGCACCGCGCAGCAGAGCGCGCGCTTCCAGATTGGAGAC 1783
:
: RESULT 2
: US-09-972-758-1
: Sequence 1, Application US/09972758
: Patent No. US20020160497A1
: GENERAL INFORMATION:
: APPLICANT: Case Western Reserve University
: APPLICANT: Montano, Monica
: APPLICANT: Wiltman, Bryan
: TITLE OF INVENTION: Suppressors of Human Breast Cancer Cell Growth
: FILE REFERENCE: 27708/04004
: CURRENT APPLICATION NUMBER: US/09/972,758
: CURRENT FILING DATE: 2001-10-05
: PRIOR APPLICATION NUMBER: US 60/238,187
: PRIOR FILING DATE: 2000-10-05
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 1
: LENGTH: 1080
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-972-758-1
:
: Alignment Scores:
: Pred. No.: 2 99e-167 Length: 1080
: Score: 1902.00 Matches: 357
: Percent Similarity: 99.72% Conservative: 1
: Best Local Similarity: 99.44% Mismatches: 1
: Query Match: 99.58% Indels: 0
: DB: 9 Gaps: 0
:
: US-09-972-758a-2 (1-359) x US-09-972-758-1 (1-1080)
:
: QY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
:   |||||||
: Db 1 ATGGCGGAGCCATTTCTTGTCAAAATATCAACACAGCTCAAACTGACATCTACAGGT 60
:
: QY 21 AlaAlaAlaValGlnGlnGlnLeuAsnProGluArgProProGlyAlaGluGluVal 40
:   |||||||
: Db 61 GCTGCTGTGCTCCAGGAAGAGCTGAACCTGAGCGCCCGCCAGGCGGAGAGCGGGTGG 120
:
: QY 41 ProGluGluAspSerArgTTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
:   |||||||
: Db 121 CCGAGAGAGACAGTGTGGTGGCATGTGAAGAGCGTTCGCCAGTGGTGGTGGTGGTGG 180
```

```
QY      61 ProGluGlyGluGlySerLeuGluSerGlnProProLeuGlnThrGlnAlaCysPro 80
Db      181 CCGGAGGGGGAGAGGAGCTCGAATCCCAACACCCTCCCTTGACAGCCAGCCGCTGTCCA 240
QY      81 GluSerSerCysLeuArgGluGlyGlyGlyGlnAsnGlyAspAspSerSerAlaGly 100
Db      241 GAATTCAGCTCCTCTGAGAGAGGCGAGAGGGCCAGAAATGGGAGCAGACTCTCTCCGTGGC 300
QY      101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuAlaGln 120
Db      301 GGCGACTTCCCGCCCGCCGAGAAAGTGGAAACCGACGCCCGAGGCCGACTCTCTCCCGAC 360
QY      121 ProCysHisAspSerSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGlu 140
Db      361 CTTGTCTAGTACTCCGAGGCCAGTAAGTTGGGGCTCTCTCCGACGAGGGCGGAAGAGAG 420
QY      141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
Db      421 TGGGACACGCGAGACAGACAGCTGGGAGAGAAAACATAGAGAGACCCCGTCCAAAGAG 480
QY      161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGlyLysLysPheAspGlu 180
Db      481 AAGCGGCAATGGAAACCGTACTACAGCTGGAAGAGAGAAAGAAAGTTTCAGAGAG 540
QY      181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
Db      541 AAACAGAGCCTTCAGACTTCAGAGATCCGAGCCGAGATGTTCCCAAGGGCCGACCGGTC 600
QY      201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
Db      601 GCGCCCTATTAACACACAGCAGATTCTCTCATGATGATACAGACCAGAGAGAGCGGATCTC 660
QY      221 LysThrGlyLeuLysSerLysArgAlaAlaLysSerAspAspThrSerAspAsp 240
Db      661 AAACCGGCTGTACTCAAGCGGCGCCGCCCAATCCACAGACACACGATGACGAG 720
QY      241 PheMetGluGluGlyGlyGlyGluAspGlyLysSerAspGlyMetGlyGlyAspGlySer 260
Db      721 TTCATGGAAACAGCGGGGTGACGAGGATGGGCGACGATGGATGGAGGGGCGGCGAC 780
QY      261 GluPheLeuGlnIleArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280
Db      781 GACTTCTGTGAGCGGAGCTTCTCGAGACCTACGACGGTACACACGGAGACCTCTCGCGC 840
QY      281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuLysCysLeuSerArg 300
Db      841 AACATGACGACGAGAGAGCTCATCAAGAGTACTGGAACGTGAGAAAGTGGCTCTCGCGC 900
QY      301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyLysAspAla 320
Db      901 ATGGACGACGAGAACACCGGCTGCGGCTGAGAGACCAACGCGCTGGGTGGCAGCACGCG 960
QY      321 ArgValArgGluLeuGluLeuLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340
Db      961 CGGTGCGGAGAGCTGAGCTGAGCTGACCGGCTGCGCGGAGAACTCCAGCTCGCTG 1020
QY      341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359
Db      1021 ACCGAAACGAACTGCACCCGACGACGAGGAGGCGCCGCTTTCACAGTTTGAGAC 1077

RESULT 3
US-09-983-965-4895
; Sequence 4895, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengding
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
```

```
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 4895
; LENGTH: 414
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 24-LIB34-008-Q1-E1-F7
US-09-983-965-4895

Alignment Scores:
Pred. No.: 1,11e-55 Length: 414
Score: 691.00 Matches: 130
Percent Similarity: 95.62% Conservative: 1
Best Local Similarity: 94.89% Mismatches: 6
Query Match: 36,18% Indels: 0
DB: 10 Gaps: 0

US-09-972-758a-2 (1-359) x US-09-983-965-4895 (1-414)

QY      155 ArgArgProSerLysLysLysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGlu 174
Db      2 AGACGGCCCTCCAGAGAGAGCGGCTTGGAACCGTACTATACCTGACCTCGGAGAGAG 61
QY      175 LysLysLysPheAspGluLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPhe 194
Db      62 AAGAAAGAGTTGATGAGAAACAGAGACCGGAGCTTCGAGAGTTTCAGAGCCGAGATGTTTC 121
QY      195 AlaLysGlyGlnProValAlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAsp 214
Db      122 GCCAAGGGCCAGCAGTTGCTCTCTATTAACACACAGCAGCTTCCTATGATGACACGAC 181
QY      215 GlnGluGluProAspLeuLysThrGlyLeuLysSerLysArgAlaAlaLysSerAsp 234
Db      182 CAGGAGGAGCGCGATCTTAAACCGGCTCTATCCCAACGCGGCGCTGCCAAATCCGAC 241
QY      235 AspThrSerAspAspPheMetGluGlyGlyGluGluGluGluGluGluGluGluGluGlu 254
Db      242 GACACACAGCAGTGAAGACTTATGAGAAACCGGCGGAGAGATGGGGCGACGCGCGG 301
QY      255 MetGlyLysAspGlySerGluPheLeuGlnIleArgAspPheSerGluThrTyrGluArgTyr 274
Db      302 ATGGAGGAGAGCGGAGGAGCTTCTGTGAGGAGGAGCTTCTGTGAGAACTATGACGCTAC 361
QY      275 HisThrGluSerLeuGluAsnMetSerLysGlnGluLeuIleLysGluTyr 291
Db      362 CACGCGGAGAGCCTTGCAACATGACGACGAGGAGAGCTCATCAAGAGTAC 412

RESULT 4
US-09-954-531-813
; Sequence 813, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; FILE REFERENCE: 689290-77
; CURRENT APPLICATION NUMBER: US/09/954,531
; CURRENT FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US/60/233,133
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,009
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,034
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,509
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,567
; PRIOR FILING DATE: 2000-09-22
```

US-09-918-995-403					
Alignment Scores:					
Pred. NO.:	6, 21e-49	Length:	495		
Score:	619.00	Matches:	121		
Percent Similarity:	98.37%	Conservative:	0		
Best Local Similarity:	98.37%	Mismatches:	2		
Query Match:	32.41%	Indels:	0		
DB:	9	Gaps:	0		
US-09-972-756A-2 (1-359) x US-09-918-995-403 (1-495)					
OY	237	SerAspAspSphemeTclnGlutylglutlunspgllySerAspGlymetGly	256		
Db	495	ACCGATAGCAGCATTCATGTCAAGAACGGCTGAGAGCATGGGGCCACCATTGGATGGGA	436		
OY	257	GLVAspGLYSerclupheulelnArqasPheSerGIurhTYrGIuaqTYrhIsthr	276		
Db	435	GGGCGAGCAGCGAGTTTTCGACGGGACTTCTCGAGAGAGCTGCAGCGGTACCAACAG	376		
OY	277	GLSerleudinasmetserylsglnlutluelllysGIurTYreulelnleulLy	296		
Db	375	GAGCGCTTGCAAGATGAGCAAGCAGAGCTCATCAGAGTAGTACCTGGAACTGGAGAAG	316		
OY	297	CysLeuSerArgmetLunspglunsnArqlenArqlenugiserLYsarqlenGly	316		
Db	315	TGCTCTCGGCMYGGAGGACGAAACAACGGCTGGCGCTGGAGAGCAACGGCTGGGT	256		
OY	317	GLVAspAlaArgValArgIuleuGluIueuGluIueuAsParqlenArqAlaGluAsn	336		
Db	255	GGCAGCAGCCGGCTGCCGGGAGCTGGAGCTGGAGCTGGACCGGCTGCCGCCGAGAAC	196		
OY	337	LeuGlnleuLeuthrglunsnGluLeuHtsArgInngInguArqAlaProLeuSerLYs	356		
Db	195	CTCAGCTGGCTGACCGAAGCAACTGCACCGGACGAGAGCGAGCGCCCTTTCCAAG	136		
OY	357	PheGlyasp 359			
Db	135	TTTGAGAC 127			
RESULT 6					
US-09-991-936-1712					
; Sequence 1712, Application US/09991936					
; Publication No. US20030073827A1					
GENERAL INFORMATION:					
APPLICANT: Brandt, Kevin S.					
APPLICANT: Gaines, Patrick J.					
APPLICANT: Stinchomd, Dan T.					
APPLICANT: Wisniewski, Nancy					
TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE					
TITLE OF INVENTION: NUCLEIC ACID MOLECULES, PROTEINS AND USES THEREOF					
FILE REFERENCE: FC-6-01					
CURRENT APPLICATION NUMBER: US/09/991,936					
CURRENT FILING DATE: 2001-11-21					
PRIOR APPLICATION NUMBER: US/09/543,668					
PRIOR FILING DATE: 2000-04-07					
PRIOR APPLICATION NUMBER: 60/128,704					
PRIOR FILING DATE: 1999-04-09					
NUMBER OF SEQ ID NOS: 1959					
SOFTWARE: PatentIn Ver. 2.1					
SEQ ID NO 1712					
LENGTH: 549					
TYPE: DNA					
ORGANISM: Ctenocephalides felis					
US-09-991-936-1712					
Alignment Scores:					
Pred. NO.:	1, 23e-08	Length:	549		
Score:	184.00	Matches:	66		
Percent Similarity:	42.29%	Conservative:	30		
Best Local Similarity:	29.07%	Mismatches:	53		
Query Match:	9.63%	Indels:	78		
DB:	9	Gaps:	12		

US-09-972-758a-2 (1-359) x US-09-991-936-1712 (1-545)

```

QY      34  ProGlyValaGluGluArgValProGluGluAspSerArgTrpGlnSerArgAlaPhe--- 52
      |||  |||||  |||
      33  CCAAGCNGGTGAAGAAATAGT-----AGTAAATATCAATCAAAAGTGTTTTAA 80
QY      53  ---ProGlnLeuGlyGlyArgProGlyProGluGlyGlnGlySerLeuGlnSerGlnPro 71
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      81  TTAACTACTGTTGAACATCAAAATGAGTGAATAATTGAA--ATATTGATCGAATAAT 137
Db      72  Pro-----ProLeuGlnTrpGlnAla 78
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      138  CCAAAATAATACACCCGACGTACAGCGTCCAAATGTCGCCGCTTCAACACAGCGCT 197
QY      79  CysProGluSerSerCysLeuArgGluGlyGlnGlyGlnGlyGlnGlyGlnGlyGln 98
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      198  ---ACTGAAGCGCTCTATWCAGAGAAAAAGGAATGATAGTGTAT----- 239
QY      99  AlaGlyGlyAspPheProProProAlaGluValGluProTrpProGluAlaGluLeuLeu 118
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      239  ----- 239
Db      119  AlaGlnProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGlu 138
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      240  ---AGAAAGAGCAACCAACAGTGGGT-----GCAACAAATGGGAAG 278
QY      139  GluGluTrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgProSer 158
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      279  ---AAGAAAAACATAGACGAGTAAATGT 305
QY      159  LysLysLysArgHisTrpLysProGlyTrpLysLeuTrpGluGlnLysLysLysPhe 178
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      306  AAGAGAAAACCTAAT--AAACCGTACATAAACAGCCTGGACCCCAACGTAAGAATGTC 362
Db      179  AspGlyLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGln 198
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      363  ---CAAAAGAAATGATGAGAGTGTGAGTGCCTGCTAAATATATAGCCATGGGAC 419
QY      199  ProValAlaProGlyAsnThrThrGlnPheLeuMetAspPheHisAspGlnGluPro 218
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      420  ACATAGTTCCTGTGACACGAAACCAATTTCTCATGGAAGATCATGAT-- 467
Db      219  AspLeuLysThrGlyLeuTyrSerLysArgAlaAlaLysSerAspPheThr----- 236
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      468  -----GTCTCTACCAAGATTCATCTCGAGACTCGGACTTACTTAGTGT 512
QY      237  ---SerAspAspPheMet 242
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      513  CGTTCGAGACACCTCTNATG 533
Db

```

RESULT 7  
US-09-974-300-7891  
Sequence 7891, Application US/09974300  
Patent No. US20020146721A1

GENERAL INFORMATION:

APPLICANT: Berka, Randy M.

APPLICANT: Clausen, Ib Groth

TITLE OF INVENTION: Methods for Monitoring Multiple Gene

FILE REFERENCE: 10085.500-US

CURRENT APPLICATION NUMBER: US-09/974, 300

PRIOR APPLICATION NUMBER: 09/680, 598

PRIOR FILING DATE: 2000-10-06

PRIOR APPLICATION NUMBER: 60/279, 526

PRIOR FILING DATE: 2001-03-27

NUMBER OF SEQ ID NOS: 8481

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 7891

LENGTH: 2421

TYPE: DNA

ORGANISM: Bacillus clausii

US-09-974-300-7891

Alignment Scores:

Pred. No.:	6.05e-05	Length:	2421
Score:	152.00	Matches:	88
Percent Similarity:	34.15%	Conservative:	52
Best Local Similarity:	21.46%	Mismatches:	110
Query Match:	7.96%	Indels:	161
DB:	10	Gaps:	16

US-09-972-758a-2 (1-359) x US-09-974-300-7891 (1-2421)

```

QY      27  GluLeuAsnProGluArgProProGlyAlaGluGluArgValProGluGluAspSerArg 46
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      523  AAGTTAAACCTGAGAAATTTGATTCAGAGCGCCGCCACCGAGAA----- 570
QY      47  TrpGlnSerArgAlaPheProGlnLeuGlyGlyArgGProGlyProGluGlyGlnGlySer 66
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      571  ---CCTGAGCCAGGCGATGGGACGAGATGGCGCGTACGCGGAC 612
Db      571  -----CCTGAGCCAGGCGATGGGACGAGATGGCGCGTACGCGGAC 612
QY      67  LeuGlnSerGlnProProProLeuGlnTrpGlnAlaCysProGluSerSerCysLeuArg 86
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      613  GATAATGACACGGCTACACCA-----CCAGAACACAGA----- 645
QY      87  GluGlyGlnGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 105
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      646  ---GAGCGCGATGACGATGCAACCGGTGAGAGGAAACCCGCCCA 687
QY      107  AlaGluValGluProTrpProGluAlaGluLeuLeuAlaGlnProCysHisAspSerGln 126
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      688  GAGCCACACAGAGCT---CCTGAAGATGAA-----GATAGCGAT 723
QY      127  AlaSerLysLeuGlyAla-----ProAla 134
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      724  GAGAGCGCGCAATGGCGGTGAGATGGCGACGAGCGGAGATCTGACCAAGACCGCCA 783
Db      135  AlaGlyGlyGlyGluGluGluGlnGlnGlnArgGlnLeuGlyLysLysHisArg 154
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      784  GCAAGAGATGGCGACAAAGATGGAGACAGAGAGAAACCGTCTCC----- 828
QY      155  ArgArgProSerLysLysArgHisTrpLysProGlyTrp----- 168
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      829  ---TTTAACCCCAATTCATGCCACACTGTGACCTTTT 861
QY      169  ---LysLeuTrpTrpGluGlnLysLysLysPheAspGlnGlnSer 183
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      862  ACTCGGAGAGACTGCAATCGTTTGGGAGAA----- 894
Db      184  LeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProValAlaProTyr 203
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      894  ----- 894
QY      204  AsnThrThrGlnPheLeuMetAspPheHisAspGlnGluGluProAspLeuLysThrGly 223
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      895  ---GCAATGGACCTCTGATGATGAGAGCAAC--GTTAAGAAGTGGC 938
QY      224  LeuTyrSerLysArgAlaAlaAlaLysSerAsp----- 234
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      939  ACAATGCAAAATTCATGACCATGCAAGAGTACCGAGCGCTTCTCGAATATGCGAAA 998
QY      235  ---AspThrSerAspAspPheMetGlnGluGlyGlyGlnGluLysArgLysArg 253
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      999  ACTGGAAAGCTTATGCAACGCTTTGGAAAGATGATGAGAAACTGGAACGCAAGTGAT 1058
QY      254  GlyMetGlyGlyAspGlySerGlnPheLeuGlnArgAspPheSerGlnTrpGlyArg 273
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      1059  TGATCCAGTATCGAGAAATTCACAAAGCAAGACCGTTCACAAACGTAATAGGCGC 1118
Db      274  TyrHisIthr-----GlnSer-LeuGlnHisMetSerLysGlnGlnLeuLeuLysGln 290
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      1119  ATCTCCGTCCTTATTAAAGTTCGATTCAGTTCACAAATGTACGGAAGACTGTTACTTA 1178
QY      290  uTyrLeuGlnLeuGlnLysCys----- 297

```





```

: Publication No. US20030104529A1
: GENERAL INFORMATION:
: APPLICANT: Tang, Y. Tom
: APPLICANT: Liu, Chenghua
: APPLICANT: Asundi, Vinod
: APPLICANT: Zhang, Jie
: APPLICANT: Ren, Feiyuan
: APPLICANT: Chen, Rui-hong
: APPLICANT: Zhao, Qing A.
: APPLICANT: Wehrman, Tom
: APPLICANT: Xue, Aidong J.
: APPLICANT: Yang, Yonghong
: APPLICANT: Zhou, Jian-Rui
: APPLICANT: Zhou, Ping
: APPLICANT: Ma, Yungqing
: APPLICANT: Wang, Dunrui
: APPLICANT: Wang, Zhiwei
: APPLICANT: Tillinghast, John
: APPLICANT: Drmanac, Radoje T.
: TITLE OF INVENTION: NO. US20030104529A1el Nucleic Acids and
: FILE REFERENCE: Polypeptides
: FILE REFERENCE: 784CIP2B
: CURRENT APPLICATION NUMBER: US/10/037,270
: CURRENT FILING DATE: 2002-01-04
: PRIOR APPLICATION NUMBER: 09/552,317
: PRIOR FILING DATE: 2000-04-25
: PRIOR APPLICATION NUMBER: 09/488,725
: PRIOR FILING DATE: 2000-01-21
: NUMBER OF SEQ ID NOS: 1104
: SOFTWARE: pt_genes Version 1.0
: SEQ ID NO 248
: LENGTH: 7453
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)..(4362)
: US-10-037-270-248

Alignment Scores:
Pred. No.: 0.000629 Length: 7453
Score: 147.00 Matches: 103
Percent Similarity: 36.84% Conservative: 51
Best Local Similarity: 24.64% Mismatches: 170
Query Match: 7.70% Indels: 95
DB: 9 Gaps: 15

US-09-972-758a-2 (1-359) x US-10-037-270-248 (1-7453)
QY 15 ThrSerAsnCysThrGlyAlaAlaValGlnGlu-----GluLeuAsnProGluArg 32
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3948 ACTACAGATACACAGAGAGCTTCTTCAGAGAGACACGCCCAAACTAAACCTGACGAC 4007
QY 33 ProProGlyAlaGluGluArgValProGluGluAspSerArgTPrGlnSerArgAlaPhe 52
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4008 TCGGATCCGCGACGTGGAAGAGAGAGAACACGCTCTCA----- 4046
QY 53 ProGluLeuGlyArgProGluProGluGluGluGluGluGluGluGluGluGluGlu 72
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4047 GGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4106
QY 73 ProLeuGlnThrGlnAlaCysProGluSerSerCysLeuArgGluGluGluGluGlu 89
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4107 GCAGTCCAGTGTGCTGATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4166
QY 90 LysGlyGlnAsnGlyAspSerSerArgGlyGlyAspPheProProAlaGluVal 109
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4167 TCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4220
QY 110 GluProThrProGluAlaGluLeuLeuAlaGluProCysHisAspSerGluAlaSerLys 129
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4221 ---CCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4277

```

```

QY 130 LeuGlyAla-----ProAlaAlaGlyGluGluGluTrp 141
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4278 GCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4337
QY 142 GlyGlnGlnArgGlu----- 147
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4338 GGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4397
QY 148 -----LeuGlyLysLysHisArgArgProSerLysLysLys 161
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4398 GCTATGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4457
QY 162 ArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAsp---Glu 180
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4458 TGTCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4516
QY 181 LysGlnSerLeuArgAlaSer----- 187
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4517 AACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4576
QY 188 -----ArgLLeuArgAlaGluMetPheAlaLysGluProValAlaProTyrAsn 204
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4577 AACGTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4636
QY 205 ThrThrGlnPheLeuMetLysAspHisAspGlnGluGluProAspLeuLysThrGlyLeu 224
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4637 ACCCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4696
QY 225 TyrSerLysArgAlaAlaAlaLysSerAsp-----AspThrSerAsp 239
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4697 GAGCTCAACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4756
QY 240 --AspPheMetGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 259
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4757 CAGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4816
QY 259 LysSerGluPheLeuGlnArgAspPheSerGluThrGlyGluArgTrpHisThrGlnSer 278
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4817 GAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4876
QY 279 LeuGln-----AsnMetSerLysGlnGluGluGluGluGluGluGluGluGlu 289
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4877 CTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4936
QY 290 GluTrpLeuGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 309
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4937 -----CAGCTCCGCAAGCTTCAGAGCTCAGATGAGATTAACCAAGCT-----GAA 4981
QY 310 LeuGluSerLysArgLeuGluGlyLysAsp-----AlaArgValArgGluGluGluGlu 327
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4982 TTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5041
QY 328 GluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuThrGluAsnGluLeu----- 345
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5042 AAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 345
QY 346 -----HisArgGlnGlnGluGluGluGluGluGluGluGluGluGluGluGluGlu 355
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5096 TCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5143

```

```

RESULT 10
US-10-037-270-249
: Sequence 249, Application US/10037270
: Publication No. US20030104529A1
: GENERAL INFORMATION:
: APPLICANT: Tang, Y. Tom
: APPLICANT: Liu, Chenghua
: APPLICANT: Asundi, Vinod
: APPLICANT: Zhang, Jie
: APPLICANT: Ren, Feiyuan
: APPLICANT: Chen, Rui-hong
: APPLICANT: Zhao, Qing A.
: APPLICANT: Wehrman, Tom

```



PRIOR APPLICATION NUMBER: US 60/292,336  
 PRIOR FILING DATE: 2001-05-22  
 PRIOR APPLICATION NUMBER: US 60/295,798  
 PRIOR FILING DATE: 2001-06-06  
 PRIOR APPLICATION NUMBER: US 60/297,457  
 PRIOR FILING DATE: 2001-06-13  
 PRIOR APPLICATION NUMBER: US 60/298,884  
 PRIOR FILING DATE: 2001-06-19  
 PRIOR APPLICATION NUMBER: US 60/303,459  
 PRIOR FILING DATE: 2001-07-09  
 NUMBER OF SEQ ID NOS: 1740  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO: 1666  
 LENGTH: 4301  
 TYPE: DNA  
 ORGANISM: Rattus norvegicus  
 FEATURE:  
 OTHER INFORMATION: Genbank Accession No. US20020119462A1 NM\_019206  
 US-09-917-800A-1666

Alignment Scores:  
 Pred. No.: 0.000516 Length: 4301  
 Score: 145.00 Matches: 96  
 Percent Similarity: 37.94% Conservative: 55  
 Best Local Similarity: 24.12% Mismatches: 160  
 Query Match: 7.59% Indels: 87  
 Gaps: 20

US-09-972-758a-2 (1-359) x US-09-917-800A-1666 (1-4301)

2 AlaLupProheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGlyAla 21  
 1028 TCAGAACCAAGACGAGCTGAG-----CACCAAGACGAGCTGACCTGAGCAGATGCA 1081  
 22 AlaAlaVal-----GlnGluGluLeu-----AsnProGluLys 32  
 1082 CAGACGATTGACAAAGAAATCAACGCCAAGAAATTTATGACAGTGGAGCTGAAGAA 1141  
 33 ProProGluLysAlaGluGluArgValProGluGluAspSerArgTrpGlnSerArgAlaPhe 52  
 1142 CCTGGAGCGGCGACCAAGACAGAGTGGAGAAATGAGACAGCAGCAGCTGCC--- 1198  
 53 ProGlnLeuGlyLysArgProGluGluGlnGlySerLeuGlnSerGlnProPro 72  
 1199 TCCACAGAGAGAGCCACAGCGGATCCGCTCGAGCAGAGATCGAGACTACGCCAGGTCCA 1258  
 73 ProLeuGlnThrGlnAlaCysProGluSerSerCysLeuArgGlnGlyLysGln 92  
 1259 -----AGAGCAGCTCAAGCAGATGAA 1279  
 93 AsnGlyAspAspSerSerAlaGlyLysAspPheProProAlaGluValGluProThr 112  
 1280 GAAGGAGGTGAATAATGAGTTGAGAACTGCCCGCAACAGCGAAGAGAGCATGAA 1339  
 113 ProGlu-----AlaGluLeuLeuAlaGlnPro-----CysHisAsp 124  
 1340 GCAGAGATGAGAGACAGCAGCAAGAAAGAACTGCTGGACCCAGACTTTGTGCCAA 1399  
 125 SerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGlnGluTrpLysGlnGln 144  
 1400 GCAGAA-----GGAAGACCTTGAGCTGCCATGAGAGAGCTCAGCGCAGAAA 1447  
 145 Gln-ArgGlnLeuGlyLysLysHisArgArgArgProSerLysLysLysArgHisIst 164  
 1448 CAGCGGTGATCTGTGACAAAGAA-----CGTATTGCCCTTAAACAAGAAAGAGAGACTCT 1504  
 164 PysProGlyTyrLysLeuThrTrpGlu-----GluLysLysLysPheAspGluLysGln 183  
 1505 CCGAGACCGAGAGCGCCCTGCGAGATGAGAGAGCAGCAGTACAGAGAGAGCATCA 1564  
 183 rLeuArgAlaSerArgIleArgAlaGluMetPheLysGlnProValAlaProTy 203  
 1565 GCTGATGAAGCAGCAGCTTAAGGACCACTACTCTGCGAGCGGCGCATGAC----- 1613

203 rAsnThrThrGlnPheLeuMetAspHisAspGlnGluProAspLeuLysThrGln 223  
 1614 -----CTGCTGGCAAGCCAGGAAGAGCGGAGCAGATGACAGCCG-- 1655  
 223 yLeuTyrSerLysArg-----AlaAlaAlaLysSerAspAspThrSerAspAspPh 241  
 1656 -----TACAAACACCGTATGATGAGCAGCTGAAGGTCAAGACAGCAGCAAGCAAGCGCG 1711  
 241 eMetGlnGlyGlyGlnGluGlnPyl-----GlySerAs 253  
 1712 GCTAACCAAGATCCAGAGCAGGAGAGCGCAAGACCGCATGGCATGTACAAAGACCT 1771  
 253 pGlyMetGlyGlyAspGlySerGluPheLeuGlnArgAsp-----PheSerGln 269  
 1772 GCACATCAATGTGGCGGAGCTGCCAGCAGCGGAGAGAGTCAACAGCTTCCCA 1831  
 269 uThrTyrGlu-----ArgTyrHisThrGlnSerLeuGlnAsnMetSerLysGlnGlu----- 286  
 1832 GCAGGAAGAGAGAGGAGCAGAGCGGAGAGGCTGCGAGCAGCAGAGAAACAGAGACCA 1891  
 287 -----LeuIleLys-----GluTyrLeuGlnLe 294  
 1892 GATGCGAGACATGTGTGCGACAGTGGAGAGCAACATGAACGAGCTGCGACAGCTCCAGAA 1951  
 294 uGluLysCysLeuSerArgMetGluAspGluAsnAsnArgLeuArgLeuSerLysArg 314  
 1952 TGAAGAAGTTCATCTGTTAGTGCATGCAACCAACCCAGAGCTGAAG----- 1997  
 314 gLeuGlyGlyAspAspAlaArgValArgGluLeuGlnGluLeuAspArgLeuArgAla 334  
 1998 -----GCCCTGAGACGAGAGGCCATACCAAGAGCTTGAAGGAATGGCAGACAGCTTCGGCC 2053  
 334 aGluAsnLeuGlnLeuLeuThrGlu-----AsnGluLeuHisArgGlnGlnGln 350  
 2054 ACCGAAAAAGGCCCTGGAAGAGATTGAACCAAGAAAGCGGAAACAGAA 2105

RESULT 12  
 US-09-291-417-106  
 Sequence 106 Application US/09291417A  
 Publication No. US20030050230A1  
 GENERAL INFORMATION:  
 APPLICANT: PLOMAN, GREGORY  
 APPLICANT: MARTINEZ, RICARDO  
 APPLICANT: WHITE, DAVID  
 TITLE OF INVENTION: STEP20-RELATED PROTEIN KINASES  
 FILE REFERENCE: 240/300  
 CURRENT APPLICATION NUMBER: US/09/291,417A  
 CURRENT FILING DATE: 1999-04-13  
 EARLIER APPLICATION NUMBER: US 60/081,784  
 EARLIER FILING DATE: 1998-04-14  
 NUMBER OF SEQ ID NOS: 147  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO: 106  
 LENGTH: 2962  
 TYPE: DNA  
 ORGANISM: Full Length Mammalian (Human) GEX2  
 US-09-291-417-106

Alignment Scores:  
 Pred. No.: 0.000792 Length: 2962  
 Score: 141.00 Matches: 103  
 Percent Similarity: 39.65% Conservative: 56  
 Best Local Similarity: 25.69% Mismatches: 145  
 Query Match: 7.38% Indels: 98  
 Gaps: 23

US-09-972-758a-2 (1-359) x US-09-291-417-106 (1-2962)  
 8 GluTyrGlnHisGlnProGlnThrSerAsnCysThrGlyAlaAlaAla----- 23  
 1784 GAGCATGGAACCAAGCAGCTGAGTACAA--GCATGAGCTGCACTGGAGCAAAATGCA 1842

```

QY 24 -----ValGInGluGluLeu-----AsnProGluArgPro 33
DB 1843 TAAAGCTTTTGACAGAAATATCAGCCAGAGAAAGTTCTTGACAGGAATTAGACAA 1902
QY 34 ProGlyAlaGluArgValProGlu---GluAspSerArgTyrPheGlnSerArgLaphe 52
DB 1903 CTTGGAGCGCTGACAGAAAGCAGCAAGTGGAGCAAGCAAGCAATGCCGTGCG--- 1959
QY 53 ProGlnLeuGlyArgProGlyProGlnGlyGluGlySer---LeuGlnSerGlnPro 71
DB 1960 CCGCCCGGAGAGAGCGCGGATCGCTGAGAGAGATCGGAGTACACAGAGTTCCA 2019
QY 72 ProProLeuGlnThrGlnAlaGlySerProGlnSerSerCysLeuArgGluGlyGly 91
DB 2020 AGAGCAGCTCAAACT-----GATCAAGAAAGAGGT 2049
QY 92 GlnAsnGlyAspAspSerSerAlaGlyGlyAspPheProProAlaGluValGlnPro 111
DB 2050 GAA-----GAGCAGGTGGAGAGCTCCCCAGACGA---GCGGAGAGA 2091
QY 112 ThrProGlnAlaGlu-----LeuLeuAlaGlnPro----- 121
DB 2092 AAGCATGAACAGACAGATGGAGAGCAGCAGCAAAAGACCTTCTTGACCGGAGCTT 2151
QY 122 CysHisAspSerGlnAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluTyr 141
DB 2152 TGTAGCCCAACAGAA-----GGAGACCTGGAGCTGGCATGAGAGGCTCAC 2199
QY 142 GlyGlnGlnGln-ArgGlnLeuGlyLysLysHisAspArgArgProSerLysLys 161
DB 2200 CACCGACACACAGCGGAGATCTGTGACAAAGAG---CGGAGTGGCTCATGAAGAAGA 2256
QY 161 SarGlnHisTyrLysProTyrTyrLysLeuThrTyrGlu---GluLysLysLysPheAsp 180
DB 2257 GAGAGCTCTTCGAGACCGGAGACGACCTGTGGAGATGAGAGACGACGCTGAGAGA 2316
QY 180 uLysGlnSerLeuArgLaseArgLLeuArgAlaGluMetPheAlaLysGlyGlnPro 200
DB 2317 GAGGACACAGCTGTGAGACAGCAGCTCAACACAGCTACTTCTTCACGGGACGAG-- 2374
QY 200 LAlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluProAsp 220
DB 2375 -----CTGCTCCGCAACGATGAGAGAGGAGCGGAGCAGCAT 2409
QY 220 uLysThrGlyLeuTyrSerLysArg-----AlaAlaAlaLysSerAspSerPhe 238
DB 2410 GCAGCGC-----TACACCAAGCGCATGATAGACACAGCTGAAGGTGCGGACAGCA 2463
QY 238 PASpAspPheMetGluGluGlyGlyGluGluAspGly----- 250
DB 2464 AAAGCGCGGCTGCCCAAGATCCAGAGAGTGAAGGAGCAGCGCATGGCATGTACAA 2523
QY 251 -GlySerAspGlyMetGlyLysAspGlySerGluPheGlnArgAsp----- 266
DB 2524 GAAGAGCTCCACATCAACAGCGCGGAGCGAGCTGAGAGGAGTGAAGATCAAGCA 2583
QY 267 -PheSerGlnThrTyrGlu---ArgTyrHisThrGlnSerLeuGlnAsnMetSerLys 285
DB 2584 GTTCTCCAGAGAGAGAGAGAGAGAGAGAGTCCGAGCGCTGACAGACAGCAAGAA 2643
QY 285 nGlu-----LeuLeuLys-----GluTyr 291
DB 2644 CGAGAACCATATGGCGCATGTGGCGCATGGCAGACAACTGACGAGCTGACGA 2703
QY 291 rLeuGlnLeuGluLysCysLeuSerArgMetGluAspGlnAsnArgLeuArgLase 311
DB 2704 GCTGAGAGATGAAGAGTCCACCTCTGTGAGAGCAGCAAAACCAAGAACTGAAG-- 2758
QY 311 uSerLysArgLeuGlyLysAspAspAlaArgValArgGluLeuGlnLeuGlnLeuAsp 331
DB 2759 -----GCCCTGGATGAGAGCCATACCAAGAACTGAAGAGATGGCGGACAA 2805
QY 331 gLeuArgAlaGluAsnLeuGlnLeuLeuThrGlu---AsnGluLeuHisArgGlnGln 350

```

```

DB 2806 GCTTCGCGCCCGCAGAGAGGCTCTGAGAGAGGATCTGACACAGAGAGGAGAGGA 2865
QY 350 u 350
DB 2866 C 2866

RESULT 13
US-10-175-523-186
; Sequence 186, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
; APPLICANT: Brockman, Jeffrey
; APPLICANT: Evans, David
; APPLICANT: Hook, Derek
; APPLICANT: Klimczak, Leszek
; APPLICANT: Laeng, Pascal
; APPLICANT: Paley, Michael
; APPLICANT: Rajan, Prithi
; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
; FILE REFERENCE: 3235/J795-US3
; CURRENT APPLICATION NUMBER: US/10/175,523
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/299,151
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/317,828
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US 60/325,150
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/333,047
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 60/349,936
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/361,834
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 186
; LENGTH: 6537
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-175-523-186

Alignment Scores:
Pred. No.: 0.00298 Length: 6537
Score: 139.00 Matches: 88
Percent Similarity: 34.07% Conservative: 66
Best Local Similarity: 19.47% Mismatches: 147
Query Match: 7.28% Indels: 152
DB: 9 Gaps: 16

US-09-972-758a-2 (1-359) x US-10-175-523-186 (1-6537)
QY 7 SerGluTyrGlnHisGlnProGlnPheSerAsnGlyThrGly----- 20
DB 3442 AATAAATATTGTCATTCCTCCCGGAGACCGCCAGTCACAGGGGAGAGGAGCTGCTG 3501
QY 21 -----AlaAlaAlaValGlnGluGluLeuAsnProGluArgProGlyAla 36
DB 3502 GAGCTGCTTAAGAGAGAGAGAGAGAGATGAAGAGAGAGAGAGAGAGAGAGAGAA 3561
QY 37 GlnGluArgValProGluGluAspSerArgTyrPheGlnSerArgAlaPheProGlnLeu 56
DB 3562 GAGGAGAGAGAGAGAGAGAGAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAA 3615
QY 57 GlyArgProGlyProGluGlyGluGlySerLeuGlnSerGlnProProLeuGlnThr 76
DB 3616 -----GAGAGAGAGAGAAATATTCAAGAGCTTCCCGCAAGATTGACGAAA 3660
QY 77 GlnAlaCysProGlnSerSerCysLeuArgGlu----- 87
DB 3661 -----CCACAGTCAGTTCATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 3711

```

87 ----- 87  
Db 3712 AGGGTCGTAAACGACGAGATCAACAGCAGTGTAAACACAGACCATTTTCAGAGACG 3771  
QY 88 -----GlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGlyAspPhe 103  
Db 3772 ACAGAGTACTGATGATGAGCCCTTTCACACTGACATGACAGAGAGG----- 3816  
QY 104 ProProProAlaGluValGluProThrProGluAlaGluLeu----- 117  
Db 3817 ---CCATGCCACAGCTGAGCCTACCTGTGAGATTGAACTGGAGAGATGGCAGAGAG 3873  
QY 117 ----- 117  
Db 3874 CCAGTCTGTAGAAAGCATTTCCAGCATCAGCTGGGAGAAACACAAACAGAGAGAGAG 3933  
QY 118 -----LeuAlaGlnProCysHisAspSer----- 125  
Db 3934 GAAGAGAAAGACATATCTTCTTCACAGATGCTGACCTTGTAGAAACATATGATGATG 3993  
QY 126 -----GluAlaSerLysLeuGlnValProAlaAlaGlyGlyGluGlu 139  
Db 3994 GATTCAAGTACTTTGAAAGAGGACGTAAGACAAATCCGACCTTAAGTGCAACAA 4053  
QY 140 GluThr-GlyGlnGlnAlaArgGlnLeu-----GlyLysLysHisArgArgAr 156  
Db 4054 GTGTGGCCAAAGAGAACAAAGCCGGCTGTATCTAAGTGGAGGCAAAACAAAGAGAGAG 4113  
QY 156 gProSer-----LysLysLysArgHisTrpLysProTyrTyrLys 169  
Db 4114 ACCGATTTAACCTGAATTGTACACCCCGCCAGAAACACCCCATGGAG--CCTGACGAGCA 4172  
QY 169 sleuThrTrpGluGluLysLysLysPheAspGlu---LysGlnSerLeuArgAlaSerAr 188  
Db 4173 GGTACACGTGAAGACAGAGAGACTTCAGAGAAAGAAACAGCCCGAGTCCCATCAG 4232  
QY 188 gLleArgAlaGlnMetPheAlaLysGlyGln----- 198  
Db 4233 GATTGACGAGGAGGTCAAGAAACTGGGGAACCCCTTGTCCCTCAAGAGCAAAACAGAG 4292  
QY 199 -----ProValAlaProTyrAsnThrThrGlnPheLeuMetAspAspH 213  
Db 4293 GGAAGAAACATGTGCCCTGTAACTCA---AACACATCA-----CCAGG 4334  
QY 213 sAspGlnGluGlnProAspLeuLysThrGlyLysLysArgAlaAlaLysSe 233  
Db 4335 TGAATAACSCAGAGATGATCTC-----ATCAAACTTGAGGAGAGAGAGAGAGAG 4385  
QY 233 rAspAspThrSerAspAspPheMetGluGluGlyGlyGluLysAspGlyGlySerAs 253  
Db 4386 GGAAG 4440  
QY 253 pGlyMetGlyLysAspGlySerGluPheLeuGlnArgAspPheSerGluThrTyrGluAr 273  
Db 4441 -----GGAATGTAGAGAAAGATCCAGATGCTGCTTAAAGCCAAAGAAA 4484  
QY 273 gTyrHisThrGluSerLeuLysMetSerLysGlnGluLeuLysGluTyrGluGlu 293  
Db 4485 AGAGAGAAACAGAAATCTCCACGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4544  
QY 293 uLeuGluLysCysLeuSerArgMetGluAspGluAspAsnArgLeuArgLeuGluSerLys 313  
Db 4545 GAGAGAGAGAA-----GAGGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4599  
QY 313 sArgLeuGlyLysAspAlaArgValArgGluLeuGluLeuLysAspArgLeuAr 333  
Db 4590 TGCCGATGACGAGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4649  
QY 333 gAlaGluAsnLeuGlnLeuLeuThrGluAsnGlu 344  
Db 4650 CAGAGAAAGCTTCAAG 4683

RESULT 14

US-09-927-597-1  
: Sequence 1, Application US/09927597  
: Publication NO. US20030032018A1  
: GENERAL INFORMATION:  
: APPLICANT: Malik, Fady  
: APPLICANT: Beraud, Christophe  
: APPLICANT: Freedman, Richard  
: APPLICANT: Craven, Andrew  
: APPLICANT: Sakowicz, Roman  
: APPLICANT: Hartman, James  
: TITLE OF INVENTION: Human smooth muscle myosin heavy chain  
: FILE REFERENCE: CYTOPO18  
: CURRENT APPLICATION NUMBER: US/09/927,597  
: CURRENT FILING DATE: 2001-08-10  
: NUMBER OF SEQ. ID NOS: 17  
: SOFTWARE: FastSeq for Windows Version 4.0  
: SEQ. ID NO 1  
: LENGTH: 5835  
: TYPE: DNA  
: ORGANISM: Human  
US-09-927-597-1  
Alignment Scores:  
Pred. No.: 0.00291 Length: 5835  
Score: 138.50 Matches: 123  
Percent Similarity: 33.51% Conservative: 64  
Best Local Similarity: 22.04% Mismatches: 130  
Query Match: 7.25% Indels: 245  
Gaps: 29  
US-09-972-758a-2 (1-359) x US-09-927-597-1 (1-5835)  
QY 14 GlnThrSerLysncysThrGlyAlaAlaValGlnGluLeuAsnProGlu----- 31  
Db 1599 CAACCTCCAGAGGTGTGCTGCTGCTGCTGGA---CGAGGAATGCTGCTTCCCAAGCCAC 1655  
QY 32 -----ArgProProGlyAlaGluGlu 38  
Db 1656 GGACAGTCTTTCTGTGAGAGCTGTGCACGGAGCAGGAGCAGCCACC----- 1703  
QY 39 ArgValProGluLysSerArg----- 46  
Db 1704 CAAGTTCACAGAGAGCCCAAGAGAGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1763  
QY 47 TrpGln----- 48  
Db 1764 TGGAGAGGTGAGACTATATGCGAGATGCTGTGACCAAGAAATATGACACCCCTGAATGA 1823  
QY 49 SerArgAlaPhePro-----GlnLeuGlyLysArg----- 58  
Db 1824 CAACGTACTTCCCTGCTCAATGAGCTCTCTCCAGAGAGATTGTGTGGCCGACCTGTGAAGA 1883  
QY 59 -----ProGlyProGluGly-----GluLysSerLeuGluSerGln 70  
Db 1884 CGTGAGACGAGATGCTGTGGCTGGAGCAGAGATGGCCAAAGATGACGAGAGAGAGAGAGAG 1943  
QY 71 ProProProLeuGlnThrGlnAlaLys-ProGluSerSerCysLeuArgGlnGlyGly 90  
Db 1944 CG-CCCTCAAGACCCAAAG 2002  
QY 90 sGlyGlnAsnGlyAspAspSer----- 98  
Db 2003 TGGGCAAGCTGATGACACACGCTACGACACACACACACACACACACACACACACACAC 2062  
QY 98 ----- 98  
Db 2063 CCAACACAG 2122  
QY 99 -----AlaGly-----GlyAspPheProProProAlaGlu--- 108  
Db 2123 GCAATGGGGGTGCTGAG 2182  
QY 109 -----ValGlnProThr-----ProGluAlaGluLeuLeuAlaGlnProCysH 123

```
Db 2183 TCACGAGATTCCGACGCTACGAGATCTGCGCGCAATCCATCCCAAGGCTTCA 2242
OY 123 saspsergluaser----- 128
Db 2243 TCGAGGGAGAGAGCGCTGCATTCTCATGATCAAAAGCCCTGGAACCTTCAACCTTAT 2302
OY 129 -----LysleuGluAlaProAlaAlaG1 136
Db 2303 ACAGATAGGCGAGAGCAAAATCTTTCGAACTGCGCTGCTGCGCCACCTAGAGAGG 2362
OY 136 Y-----GlyGluGluGluTrp----- 141
Db 2363 AGCGAGATTGAGATCACCCATGTCATCATGCTTCCAGCGCATGTGTGCTGCTACT 2422
OY 142 -GlyGluGln-----GlnArgGlnLeuGlyLysLys 152
Db 2423 TGGCCAGAAAGGCTTTTGGCAAGAGGAGAGACAGACAGCTGACCGCCATGAAGGATTCAGA 2482
OY 152 shisArgArgArgProSerLysLysArgHisTrp----- 164
Db 2483 GGAACTGGCGCG-CCTACCTC-AAGCTGCGAACTGCGAGTGTGAGAGCTTTCCACCA 2540
OY 165 ----LysProTyrTyrLysLeuThrTrpGluGluLysLys-----LysPheAspG1 180
Db 2541 AGTGAAGCCACTGCTGCGAGGTGACGCGAGAGAGAGAGATGACGCCCAAGAGAGATGA 2600
OY 180 uLysGlnSerLeuArgAlaSerArg1LeArgAlaGluMetPheAlaLysGlyGlnProVa 200
Db 2601 ACTCGAAGAACCAAGAGAGCGCGACAGACAGAGCAAGCAAG----- 2637
OY 200 1AlaProTyrAsnThrThrGlnPheLeuMetAspHisAspGln-----GluGluPr 218
Db 2638 -----AATGAGCTTAAAGAGCTGAAACAGAGAGCTGCGCTGACGAGAGAGAA 2687
OY 218 oAspleu-----LysThrGlyLeuTyrSerLys----- 227
Db 2688 GAACCTGCTACAGAGACAGCTGCGAGGACAGACAGAGCTGTATGAGAGGCTGAGAGAT 2747
OY 228 -----ArgAlaAlaAlaLysSerAspAspThrSerAspAspPheMetGluGluG1 245
Db 2748 GCGGCTGCGGCTGCGGCGCCAGAGAGACAGAG-----CTGAGAGAGAT 2789
OY 245 yGlyGluLysProGlyGlySerAspGlyMetGlyGly-AspGlySerLysPheLeuGlnA 265
Db 2790 -ACTGCATGAGATGAGGCGCGCTGAGAGAGAGAGAGAGAGGCGCCAGAGCTACAGG 2848
OY 265 rgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGlnAsnMetSerLysG 285
Db 2849 CTGAAGAGAGAGATGCGCCAGCAGATG---CTGAGACCTTGAAGAAACGCTGAGAGAG 2905
OY 285 lngluLeu1LeuGlyGluTyrLeuGluLeuGluLysCysLeuSerArg----- 300
Db 2906 AGGAAGCTGCCAGCGAGAGAGCTGCAACTTGAAGAGATCACCGCTGAGGCCAAGATCAAGA 2965
OY 301 -----MetGluAspGluAsnAsnArgLeuArgLeuGluSerL 313
Db 2966 AACTGAGAGATGACATCTGCTCATGTGATTCAGAAACAAATAAATCAAAAGAGAGCA 3025
OY 313 ysArgLeuGlyGlyAspAspAlaArgValArgGluLeuGluLeuGluLeu-----AspA 331
Db 3026 AACTCGTT-----GAGGAGAGATTAAGTACTTAACGACAAATCTTCCAGAGAGG 3076
OY 331 rGluAlaGluAlaGluAsnLeu-----GlnLeuLeuThrGluA 343
Db 3077 AAGAAAGGCCAAGAAATCTTACCAAGCTGAAGAAACAGACATGATCATATGTTTCAACAAC 3136
OY 343 snGluLeuHisArgGlnGlnGlu-----ArgAlaProLeuSerLys 356
Db 3137 TGGAAATGGCGCTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3186
OY 356 -----ValGluProThr-----ProGluAlaGluLeuLeuAlaGluProCysH1 123
```

RESULT 15  
US-09-927-597-3

```
Sequence 3, Application US/09927597
Publication No. US20030032018A1
GENERAL INFORMATION:
APPLICANT: Malik, Fady
APPLICANT: Betaud, Christophe
APPLICANT: Freedman, Richard
APPLICANT: Craven, Andrew
APPLICANT: Sakowicz, Roman
APPLICANT: Hartman, James
TITLE OF INVENTION: Human smooth muscle myosin heavy chain
FILE REFERENCE: CYT0018
CURRENT APPLICATION NUMBER: US/09/927,597
CURRENT FILING DATE: 2001-08-10
NUMBER OF SEQ ID NOS: 17
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 3
LENGTH: 5937
TYPE: DNA
ORGANISM: Human
US-09-927-597-3
```

```
Alignment Scores:
Pred. No.: 0.00297 Length: 5937
Score: 138.50 Matches: 123
Percent Similarity: 33.51% Conservative: 64
Best Local Similarity: 22.04% Mismatches: 130
Query Match: 7.25% Indels: 245
DB: 9 Gaps: 29
```

US-09-972-758a-2 (1-359) x US-09-927-597-3 (1-5937)

```
OY 14 GlnThrSerAsnCysThrGlyAlaAlaAlaValGlnGluGluLeuAsnProGlu----- 31
Db 1509 CAACCTCCAGAGTGTGCTGCGCTGCGCTGCGA---CGAGAAATGCTGTGTCCCAAGGAC 1655
OY 32 -----ArgProGlyAlaGluGlu 38
Db 1656 GGACAAGCTTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1703
OY 39 ArgValProGluAspSerArg----- 46
Db 1704 CAAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1763
OY 47 TrpGln----- 48
Db 1764 TGGGAAGTGAATATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1823
OY 49 SerArgAlaPhePro-----GlnLeuGlyArg----- 58
Db 1824 CAACGTGACTTCTGCTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1883
OY 59 -----ProGlyProGlyGly-----GlnGlySerLeuGluSerGln 70
Db 1884 CHTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1943
OY 71 ProProLeuGlnGlnGlnAlaCys-ProGluSerSerCysLeuArgGluGlyGluGly 90
Db 1944 CG-CTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2002
OY 90 sGlyGlnAsnGlyAspAspSer----- 98
Db 2003 TGGGCAAGCTGATGACACAGCTACGCAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2062
OY 98 ----- 98
Db 2063 CCAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2122
OY 99 -----AlaGly-----GlyAspPheProProAlaGlu-- 108
Db 2123 GCATGGGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2182
OY 109 -----ValGluProThr-----ProGluAlaGluLeuLeuAlaGluProCysH1 123
```

```
Db      2183  TCCAGAGATTCCGCCAACGCTAGAGATCCTGGCGGGAATGCCATGCCCAAGGCTTCA 2242
QY      123  sAspSerGluAlaSer----- 128
Db      2243  TGGACGGGAAGCAGCGGCTGCTATTCATGATCAAAAGCCCTGGAACCTTGACCCCACTTAT 2302
QY      129  -----LysLeuGlyValProAlaIaG1 136
Db      2303  ACAGGATAGGCGACAGCAAAATCTTCTCCGAACGTGGCTCCGCCACCTAGAGGAGG 2362
QY      136  y-----GlyGluGluIuTrp----- 141
Db      2363  AGCGAGATTGAAGATCACCAGATGTCAATACATGCGCTTCACAGCGATGTGTGGCTACT 2422
QY      142  -GlyGluGln-----GlnArgGlnLeuGlyLysLysLys 152
Db      2423  TGGCCAGAAAGGCTTTTGGCAGAGCAGCAGCAGCTGACCGCCATGAGGTGATTCAGA 2482
QY      152  sHisArgArgArgProSerLysLysLysArgHisTrp----- 164
Db      2483  GGAACCTGCGCCG-CCTACCTC-AAAGCTGGGAAGCTGGCAGCTGTGAGACCTTTTCACCA 2540
QY      165  ---LysProTyrTyrLysLeuThrTrpGluGluLysLys-----LysPheAspG1 180
Db      2541  AGTGAAGCCACTGCTGTCAGAGTCAGACGAGCAGAGAGATGCAGCCAGAGAGATGA 2600
QY      180  uLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVa 200
Db      2601  ACTGCAGAACCAAGCAGCGCGCAGCAGAACGACAG----- 2637
QY      200  lAlaProTyrAsnThrGlnPheLeuMetAspAspHisAspGln-----GluGluPr 218
Db      2638  -----AATGAGCTTAAGAGAGCTGAGACAGAGACACTGCGACCTGACCGAGAGAA 2687
QY      218  oAspLeu-----LysThrGlyLeuTyrSerLys----- 227
Db      2688  GAACCTGCTACAGAACAGCTGCAGGCGAGAGACAGAGCTGTATGTCAGAGGCTGAGAGAT 2747
QY      228  -----ArgAlaIaAlaLysSerAspAspThrSerAspAspPheMetGluGluG1 245
Db      2748  GCGGCTGGGCTGGCGGCCAAGAACAGAGAG-----CTGGAGAGAT 2789
QY      245  yGlyGluGlnAspGlySerAspGlyMetGlyGly-AspGlySerGluPheLeuGlnA 265
Db      2790  -ACTGCATGACATGGAGGCCCGCCCTGAGAGAGAGAGACAGCGGCCACGACGTACAGG 2848
QY      265  rGAspPheSerGluThrTyrGluArgGlyrHisThrGlnSerLeuGlnAsnMetSerLysG 285
Db      2849  CTGAAGGAGAGAGATGGCCCGCAGCATG---CTGGACCTTGAAAGACAGCTGGAGGAGG 2905
QY      285  lInGluLeuIleLysGluTrpLeuGluLeuGluLysCysLeuSerArg----- 300
Db      2906  AGGAAGCTGCCAGGACGAAGCTGCACACTTGGAGAAGGTCAAGGCTGAGGCCAAGATCAAG 2965
QY      301  -----MetGluAspGluAsnAsnArgLeuArgLeuGluSerL 313
Db      2966  AACTGAGAGATGAGATCCTGTCATGATGATACAGACATATAACTATCAAAAGACGAA 3025
QY      313  ysArgLeuGlyGlyAspAspAlaArgValArgGluLeuGluLeuLeu-----AspA 331
Db      3026  AACTTCCTT-----GAGGAGAGGATTAGTACTTAACGACCAAAATCTTGCAGAGAGG 3076
QY      331  rGLeuArgAlaGluAsnLeu-----GlnLeuLeuThrGluA 343
Db      3077  AAGAAAGGCCACAGATCTTACCAAGCTGAAGAAACACAGCATGATCTATGATTTCAAGAC 3136
QY      343  snGluLeuHisArgGlnGlnGlu-----ArgAlaProLeuSerLys 356
Db      3137  TGGAGTGGCGCTAAAGAAAGAGAGAACGCCGACAGAGCTGGAGAG 3186
```

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_p1us\_p2n model

Run on: June 17, 2003, 09:52:48 ; Search time 3255 seconds  
(without alignments)  
3209.803 Million cell updates/sec

Title: US-09-972-758a-2

Perfect score: 1910  
Sequence: 1 MAEPFLSEYOHOPOTSNCCTG.....LTENELHROQERAPLSKFGD 359

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODE=frame+pzn.model -DEV=xlh  
-O=/cg2.1/uspto.spool/us09972758/runat.17062003.095237.24371/app.query.fasta.1.519  
-DB=GenEmbl -QFMT=fastop -SUFFIX=rgc -MINMATCH=0.1 -LOOFCR=0 -LOOFEEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blotsum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=us09972758@cgn1.1.2496 -runat.17062003.095237.24371 -NCPU=6 -ICPU=3  
-NO\_MMAPP -LARGEQUERY -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV.TIMEOUT=120 -MARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_htg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vl: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sts: \*  
28: em\_un: \*

29: em\_vl: \*  
30: em\_htg\_hum: \*  
31: em\_htg\_in: \*  
32: em\_htg\_om: \*  
33: em\_htg\_ov: \*  
34: em\_htg\_pat: \*  
35: em\_htg\_ph: \*  
36: em\_htg\_pl: \*  
37: em\_htg\_ro: \*  
38: em\_htg\_sts: \*  
39: em\_htgo\_hum: \*  
40: em\_htgo\_mu: \*  
41: em\_htgo\_other: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1910	100.0	2086	9 BC006460	BC006460 Homo sapi
2	1910	100.0	3624	9 AB021179	AB021179 Homo sapi
3	1910	100.0	100836	2 AC087298	AC087298 Homo sapi
4	1903	99.6	2178	9 AK023624	AK023624 Homo sapi
5	1858.5	97.3	142326	2 AC024047	AC024047 Homo sapi
6	1587.5	83.1	3402	10 AY090614	AY090614 Mus muscu
7	1587.5	83.1	111002	10 AL731805	AL731805 Mus muscu
8	1577.5	82.6	1488	10 BC022111	BC022111 Mus muscu
9	1514.5	79.3	175120	2 AC120950	AC120950 Rattus no
10	629	32.9	461	6 AX331371	AX331371 Sequence
11	556.5	29.1	1330	9 AK056946	AK056946 Homo sapi
12	550	28.8	116035	2 AC008105	AC008105 Homo sapi
13	530	28.8	223442	10 AL662804	AL662804 Mouse DNA
14	545.5	28.6	1238	9 BC025970	BC025970 Homo sapi
15	545	28.5	2048	9 AK096785	AK096785 Homo sapi
16	214	11.2	1536	3 AY051786	AY051786 Drosophill
17	205.5	10.8	183439	3 AC007809	AC007809 Drosophill
18	205.5	10.8	232744	3 AE003705	AE003705 Drosophill
19	203.5	10.8	237119	2 AC017740	AC017740 Drosophill
20	187	9.8	165197	2 AC112538	AC112538 Rattus no
21	183	9.6	59599	2 AC107153	AC107153 Rattus no
22	170	8.9	9121	4 OATRICHA	Z18361 O.aries tri
23	165.5	8.7	9344	4 OATRICHA	Z18361 O.aries tri
24	165	8.6	6644	4 RABMACP	M77812 Rabbit myos
25	164.5	8.6	4143	5 GSDYNACT	X62773 Gallus gall
26	162	8.5	4152	9 HUMMYH9	M81105 Homo sapien
27	162	8.5	278652	2 AC073809	AC073809 Mus muscu
28	160	8.4	241928	2 AC123859	AC123859 Mus muscu
29	159	8.3	151578	9 AL589986	AL589986 Human DNA
30	159	8.3	256346	2 AC105970	AC105970 Mus muscu
31	158.5	8.3	9551	6 AR076233	AR076233 Sequence
32	158.5	8.3	9551	6 I39845	I39845 Sequence 93
33	158.5	8.3	9551	6 HUMTRHVAL	L09190 Human trich
34	158.5	8.3	126357	2 AC121480	AC121480 Rattus no
35	157.5	8.2	74317	8 NC12F11	AL451017 Neurospor
36	157	8.2	174241	9 AC018634	AC018634 Human Chr
37	156	8.2	170778	8 OSJN00052	AL606627 Oryza sat
38	156	8.2	174019	2 AC027267	AC027267 Homo sapi
39	155	8.2	185155	2 AC073446	AC073446 Homo sapi
40	155.5	8.1	5122	9 HUMMYNM	M31013 Human nomu
41	155.5	8.1	174472	2 AC074356	AC074356 Mus muscu
42	155	8.1	2277	6 AR004983	AR004983 Sequence
43	155	8.1	5002	10 DB9728	DB9728 Mus musculu
44	155	8.1	143487	2 AC016033	AC016033 Homo sapi
45	155	8.1	173934	2 AC100756	AC100756 Homo sapi

RESULT 1

ALIGNMENTS

BC006460  
 LOCUS BC006460 2086 bp mRNA linear PRI 12-JUL-2001  
 DEFINITION Homo sapiens, HMBA-inducible, clone MGC:1880 IMAGE:3535529, mRNA,  
 complete cds.  
 ACCESSION BC006460  
 VERSION BC006460.1 GI:13623668  
 KEYWORDS MGC.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 2086)  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-APR-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 COMMENT Contact: MGC help desk  
 Email: [cgapsb-email.nih.gov](mailto:cgapsb-email.nih.gov)  
 Tissue Procurement: DCTD/DTP  
 CDNA Library Preparation: Rudin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Genome Sequence Centre,  
 BC Cancer Agency, Vancouver, BC, Canada  
 info@cgsc.bc.ca  
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
 Susana Chan, Readman Chiu, Chris Fjell, Erin Galland, Ran Guin,  
 Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo  
 Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeay, Steven  
 Ness, Pawan Pandoh, Anna-Liisa Prabh, Parvaneh Saeedi, Jacqueline  
 Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Scott,  
 Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,  
 George Yang, Scott Zuyderduyn, Marco Marra.  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
 Series: IRAL Plate: 8 Row: d Column: 1  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 4062855.  
 FEATURES  
 Source  
 1. 2086  
 /organism="Homo sapiens"  
 /db\_xref="locusid:10614"  
 /db\_xref="taxon:9606"  
 /clone="MGC:1880 IMAGE:3535529"  
 /tissue\_type="Lung, small cell carcinoma"  
 /clone\_1ib="NH\_MGC\_7"  
 /lab\_host="DH10B-R"  
 /note="vector: POTB7"  
 606..1685  
 /codon\_start=1  
 /product="HMBA-inducible"  
 /protein\_id="AAH06460.1"  
 /db\_xref="GI:13623669"  
 /translation="MAEPLSEYOHOPOTNCTGAAYOEELNPPPCAEERPEED  
 SRWQSRAPOLGRRPGEGSLLEQPPLOTQACPESSCLRGEGGSDSSAGD  
 FPPRAVEPTPEELLAOQCHDEFAKLPAAGGEENGGQOROLGKKHRRBPBK  
 KRHKPYKLTWEKKKFKDEKSLRASRIAEKFAQVPAPNTYQFLMDHDOEP  
 DLKTGLYSKRAAKSDTSDDDFMEEGEGSDMGDSFFLORDSEYERYHT  
 ESTLONSKOELIKLEYLEKLCISRMEDENNRLRLESKRLGCGDARVLELELDRLRA  
 ENLQLTENRHLHROERAPLSKRGD"  
 BASE COUNT 510 a 505 c 618 g 453 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 6, 06e-93 Length: 2086  
 Score: 1910.00 Matches: 359  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: Gaps: 0

US-09-972-758a-2 (1-359) x BC006460 (1-2086)  
 QY 1 MetAlaGluProPheLeuSerGluTyrGlnHISGlnProGlnThrSerAsnCyThrGly 20  
 Db 606 ATGGCCGAGGACATTTCTTGTCAGAAATATCAACACACACCTCAAACTGTGACACTGTACAGGT 665  
 QY 21 AlaAlaValGlnGlnGluLeuAsnProGlnAArgProGluAlaGluGluAArgVal 40  
 Db 666 GCTGCTGCTGTCCTCCAGGAAGACCTGAACCTTGAGCGCCGCCCAAGCGGAGAGCGCGGTG 725  
 QY 41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGluLeuGlnGlyLarProGly 60  
 Db 726 CCCGAGAGAGACAGTAGTGAGCAATCGAGAGCGTCCCACTGGTGGCTCCGGG 785  
 QY 61 ProGluGlyGlnGlySerLeuGlnSerGlnProProProLeuGlnThrGlnAlaCysPro 80  
 Db 786 CCGAGGGGGAGAGGAGCGCTGGAAATCCCAACCACTCCCTTGCAAGCCAGCGCTTCCA 845  
 QY 81 GluSerSerCysLeuArgGlnGlnGlyGlyGlnGlnAsnGlyAspAspSerAlaGly 100  
 Db 846 GAATCTAGCTGCTTGAGAGAGCGGAGAGGCGCAATGAGGAGCACTGCTCCGCTGGC 905  
 QY 101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120  
 Db 906 GGGCACTCCCGCCCGCGGAGAAAGTGGAGCCGAGCGCCGAGGCGGAGCTGCTCGCCAG 965  
 QY 121 ProCysHisAspSerGlnAlaSerLysLeuGlnGlyAlaProAlaAlaGlyGlnGlnGlu 140  
 Db 966 CCTTGTGATGATCTCCAGGCGCAATGAGTGGGGGCTCTCCGCGAGGGGGCGAAGAGAG 1025  
 QY 141 TrpGlnGlnGlnGlnArgGlnLeuGlnGlyLysLysHisArgArgArgProSerLysLys 160  
 Db 1026 TGGGACAGACAGACAGACAGACAGCTGGGGGAAAAAATATGAGAGAGCCCTTCAGAAAG 1085  
 QY 161 LysAlaGlnIstTrpLysProTyrTyrLysLeuThrTrpGlnGluLysLysLysPheAspGlu 180  
 Db 1086 AACCGGCAATGGAAACCGTACTCAAGCTGAGCTGGAGAGAGAAAAAGTTCGACGAG 1145  
 QY 181 LysGlnSerLeuArgAlaSerArgTgLeArgAlaGlnPheAlaLysGlnGlnProVal 200  
 Db 1146 AAACAGAGCCCTTGACCTTCAAGATCCAGGCGCAATGCTGCCAAGGCGCAAGCGGGTCT 1205  
 QY 201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluProAspLeu 220  
 Db 1206 GCGCCCTATAACACCAACGAGGATCTCTCATGATGATGATGATGATGATGATGATGATGATGAT 1265  
 QY 221 LysTrpGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAsp 240  
 Db 1266 AAACCCGCGCTGTACTCCAAAGCGGGCGCGCCCAATTCGACGACACCGCATGACGAC 1325  
 QY 241 PheMetGlnGlnGlnGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260  
 Db 1326 TTCTATGAGAGAGAGGGGGTGAAGAGATGGGGGAGCGATGGAGTGGAGAGCGGCGAC 1385  
 QY 261 GluPheLeuGlnLarGspPheSerGlnThrTyrGlnLarGlyHisThrGlnSerLeuGln 280  
 Db 1386 GATTTCTGACAGGGGACCTTCTGAGACGCTACGACCGGTATACACGAGAGACCGTCCGAG 1445  
 QY 281 AsnMetSerLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300  
 Db 1446 AACATGACGACAGACGAGCTCATCAAGAGTACCTGAGCTGAGAGAGAGGCTCTCCGCGC 1505  
 QY 301 MetGluAspGluAsnAsnArgLeuArgLeuGlnGlnSerLysArgLeuGlnGlnLysAspAla 320  
 Db 1506 ATGAGAGGACGAGAACACCGGCTGCGTGGAGAGCAACCGGCTGGTGGCACGACGCGG 1565  
 QY 321 ArgValArgGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 340  
 Db 1566 CCGTGTGCGGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCT 1625  
 QY 341 ThrGluAsnGlnLeuLeuHisArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 359

Db 1626 ACCGAGACGAACCTGCACCGGAGAGACGCGCTTCCAAAGTTGGAGAC 1682

RESULT 2  
LOCUS AB021179 3624 bp mRNA linear PRI 05-DEC-2000  
DEFINITION Homo sapiens mRNA for HEXIM1 protein, complete cds.  
ACCESSION AB021179  
VERSION AB021179.1 GI:4062855  
KEYWORDS HEXIM1 protein.  
SOURCE Homo sapiens coronary artery smooth muscle cell cDNA to mRNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Kusunohara, M., Nagasaki, K., Kimura, K., Maass, N., Manabe, T.,  
Shikawa, S., Aikawa, M., Miyazaki, K. and Yamaduchi, K.,  
Cloning of hexamethylene-bis-acetamide-inducible transcript,  
HEXIM1, in human vascular smooth muscle cells  
Biomed. Res. 20, 273-279 (1999)  
JOURNAL 2 (bases 1 to 3624)  
REFERENCE Submitted (14-DEC-1998) Masatoshi Kusunohara, National Defense  
AUTHORS Medical College, The First Department of Internal Medicine; 3-2  
TITLE Namihi, Tokorozawa, Saitama 359-8513, Japan  
JOURNAL (E-mail: mkusunoh@ndmc.ac.jp, Tel: +81-42-995-1597,  
Fax: +81-42-996-5200)

FEATURES  
source location/Qualifiers  
1..3624  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_type="coronary artery smooth muscle cell"  
1..3624  
/gene="HEXIM1"  
690..1769  
/gene="HEXIM1"  
/codon\_start=1  
/product="HEXIM1 protein"  
/protein\_id="BAA36166.1"  
/db\_xref="GI:4062855"  
/translation="MAEPFLSEYOHQPOHSNCTGAANOBEILNPPRPAGEEVRPED  
SRKOSRAFPQLGCRPGPEEGSELESPPPLQDACPSSCLREGEFGMDSDSAGCD  
FPPPAVEEPTPEAEILAOPCHOSEASKLAPAGAGEEFGQOQRLGKKRHRRPSKK  
KHHMKPYRLTWTEEKKRFDKQSLRASRIAEFMFAGOPVAPYNTQFLMDQDEP  
DIKTGLYSRAAKASDSDTDDDFMEEGEDGSDMGDSGSEFLORDFSERYHT  
ESLONMSKOELIKEYLEKLSRMEDENNRLRLFSKRLGGDABVBELELDRLRA  
ENLQLTREBLRHQDERAPLSKRGD"

BASE COUNT 978 a 775 c 964 g 907 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1..06e-92 Length: 3624  
Score: 1910.00 Matches: 359  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0

US-09-972-758a-2 (1-359) x AB021179 (1-3624)

QY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCythrGly 20  
|||||  
Db 690 ATGGCCGACCATTTCTTGTCAGAAATCAACACACACCTCAACTAGCAACTGTACAGGT 749  
|||||  
QY 21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProGluAlaGluGluArgVal 40  
|||||  
Db 750 GCTGCTGCTGTCAGGAAGAGCGTGAACCCCTGAGCGGCCCGGAGGAGCGCGTg 809  
|||||  
QY 41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60  
|||||  
Db 810 CCGGAGAGAGACAGTAGTGCGCAATGAGAGAGCTTCCCGCCAGTTGGTGCGCGT 869  
|||||  
QY 61 ProGluGlyGluGlySerLeuGlnSerGlnProProProLeuGlnThrGlnAlaCysPro 80  
|||||

Db 870 CCGAGAGGGGAACGAGCCTGGAAATCCCAACCACTCCCTTGGACAGCCAGGCTGTCCA 929  
|||||  
QY 81 GluSerSerCysLeuArgGluGluGlyGlyGlnAsnGlyAspAspSerSerAlaGly 100  
|||||  
Db 930 GAATCTAGCTGCTCTGAGAGAGGGGCGAGAAAGGGCCAGAAAGGGGACACTGCTCCGTGGC 989  
|||||  
QY 101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120  
|||||  
Db 990 GCGGACTTCCCGCGCGCGGCGAGAGTGGAAACGACGCCCGGAGGCCAGCTGCTCGCCGAG 1049  
|||||  
QY 121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGluGluGlu 140  
|||||  
Db 1050 CTTTGATGACTGCTCCGAGGCCAGTAGATTGGGGGCTCTGCGCGAGGGGCGAAGAGAG 1109  
|||||  
QY 141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysHisArgArgArgProSerLysLys 160  
|||||  
Db 1110 TGGGACACACAGCAGAGAGACGCTGGGGGAAAAAACAATAGAGAGAGCCCGTCCAAGAG 1169  
|||||  
QY 161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 180  
|||||  
Db 1170 AACCGGCAATTGGAAACCGTACTACAGCTGACCTGGGAAGAGAGAAAAAGTTCCAGCAG 1229  
|||||  
QY 181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200  
|||||  
Db 1230 AACAGAGCCTTCGAGCTTCAAGGATCCGACCGAGAGATGTTCGCCAAGGGCCAGCCGCTC 1289  
|||||  
QY 201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluProAspLeu 220  
|||||  
Db 1290 GCGCCCTATACACCCAGCAGAGTTCTCATGATGATGATCAGACACGAGAGAGCGGATTC 1349  
|||||  
QY 221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAsp 240  
|||||  
Db 1350 AAACCGGCTGTACTCCAAAGCGGGCGCGCCCAATTCGACAGCACCCAGCATGACGAC 1409  
|||||  
QY 241 PheMetGluGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyLysAspGlySer 260  
|||||  
Db 1410 TTCATGGAAGAAAGGGGCTGAGAGATGGGGGAGCATGGATGGATGGAGAGGAGCGACAC 1469  
|||||  
QY 261 GluPheLeuGlnArgAspPheSerGluTyrTrpGluArgTyrHisThrLysSerLeuGln 280  
|||||  
Db 1470 GAATTTCTCAGGGGAGCTTCTCGAGACGTACGAGCGGTACCAACGGAGAGCCCTGCAG 1529  
|||||  
QY 281 AsnMetSerLysGlnGlnLeuLeuLysGlyTyrLeuGlnLeuGlnLysCysLeuSerArg 300  
|||||  
Db 1530 AACATGACGACAGCAGAGAGCTCATCAGCACTACTGTAAGTGGAGAAGTCCCTTCGCC 1589  
|||||  
QY 301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyLysAspAla 320  
|||||  
Db 1590 ATGGAGAGCAGAGAACCAACCGGCTGCGGTGAGAGCAACCGGCTGGCGCAGCACCGC 1649  
|||||  
QY 321 ArgValArgGluLeuGlnGluLeuLeuAsnArgLeuArgAlaGlnAsnLeuGlnLeuLeu 340  
|||||  
Db 1650 CGTGTCCGGAGCTGGAGCTGGAGCTGGACCGGCTCGCGCAGAACTCCAGCTGCTG 1709  
|||||  
QY 341 ThrGluAsnGluLeuLeuLysArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359  
|||||  
Db 1710 ACCGAGACGAACACTGCACCGGCGAGCAGAGCGCGCCCTTCCAAAGTTTGGACAGC 1766  
|||||

RESULT 3  
AC087298 100836 bp DNA linear HTG 20-APR-2002  
LOCUS AC087298  
DEFINITION Homo sapiens chromosome 17 clone RP11-40363 map 17, WORKING DRAFT  
SEQUENCE 15 ordered pieces.  
ACCESSION AC087298  
VERSION AC087298.8 GI:20219131  
KEYWORDS HTG: HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 100836)  
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE Homo sapiens chromosome 17, clone RP11-403G3  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 100836)  
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Berra, N., Bastien, V., Boguslavsky, L., Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Hagos, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., Labocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Plunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Sougniez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testfaye, S., Theodore, J., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Submitted (23-DEC-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
JOURNAL 3 (bases 1 to 100836)  
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Berra, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Labocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Plunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Submitted (20-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
JOURNAL On Apr 20, 2002 this sequence version replaced gi:20177719.  
COMMENT All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
Project Information  
Center project name: LI2020  
Center clone name: 403.G.3  
----- Summary Statistics  
Sequencing vector: Plasmid; n/a: 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 96277 bases at least Q40  
Consensus quality: 98051 bases at least Q30  
Consensus quality: 98838 bases at least Q20  
Insert size: 14000; agarose-fp

Insert size: 99436; sum-of-contigs  
Quality coverage: 19.1 in Q20 bases; agarose-fp  
Quality coverage: 26.9 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 15 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
1 8098: contig of 8098 bp in length  
\* 8099 8198: gap of 100 bp  
\* 8199 8271: contig of 73 bp in length  
\* 8272 8371: gap of 100 bp  
\* 8372 9497: contig of 1126 bp in length  
\* 9498 9597: gap of 100 bp  
\* 9598 10863: contig of 1266 bp in length  
\* 10864 10963: gap of 100 bp  
\* 10964 12232: contig of 1269 bp in length  
\* 12233 12332: gap of 100 bp  
\* 12333 13844: contig of 1512 bp in length  
\* 13845 13944: gap of 100 bp  
\* 13945 15215: contig of 1271 bp in length  
\* 15216 15315: gap of 100 bp  
\* 15316 18234: contig of 2919 bp in length  
\* 18235 18334: gap of 100 bp  
\* 18335 21787: contig of 3453 bp in length  
\* 21788 21887: gap of 100 bp  
\* 21888 25485: contig of 3598 bp in length  
\* 25486 25585: gap of 100 bp  
\* 25586 40531: contig of 1496 bp in length  
\* 40532 40631: gap of 100 bp  
\* 40632 52119: contig of 11488 bp in length  
\* 52120 52219: gap of 100 bp  
\* 52220 64941: contig of 12722 bp in length  
\* 64942 65041: gap of 100 bp  
\* 65042 79587: contig of 1456 bp in length  
\* 79588 79687: gap of 100 bp  
\* 79688 100836: contig of 21149 bp in length.  
Location/Qualifiers  
1. 100836  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/map="17"  
/clone="RP11-403G3"  
/clone\_lib="RPCI-11 Human Male BAC"  
1. 8098  
/note="assembly\_fragment  
clone\_end:SP6  
vector\_side:left"  
8199..8271  
/note="assembly\_fragment"  
8372..9497  
/note="assembly\_fragment"  
9598..10863  
/note="assembly\_fragment"  
10964..12232  
/note="assembly\_fragment"  
12333..13844  
/note="assembly\_fragment"  
13945..15215  
/note="assembly\_fragment"  
15316..18234  
/note="assembly\_fragment"  
18335..21787  
/note="assembly\_fragment"  
21888..25485  
/note="assembly\_fragment"  
25586..40531

```

misc_feature      /note="assembly_firmgment"
                  40632..52119
misc_feature      /note="assembly_firmgment"
                  52220..64941
misc_feature      /note="assembly_firmgment"
                  65042..79587
misc_feature      /note="assembly_firmgment"
                  79688..100836
misc_feature      /note="assembly_firmgment"
                  79688..100836

BASE COUNT      23052 a 27082 c 27053 g 22249 t 1400 others
ORIGIN
Alignment Scores:
Pred. No.:
Score:           3e-91
Percent Similarity: 1910.00
Best local Similarity: 100.00%
Query Match:     100.00%
DB:              2
                Gaps: 0

```

Query Match: 99.63% Indels: 0  
 DB: 9 Gaps: 0  
 us-09-972-758a-2 (1-359) x AK023624 (1-2178)

1 MetAlaGluProPheUeuSerGluYrGlnHisGlnProGluThrSerAsnGlyThrGly 20  
 |||||  
 DB 716 ATGGCGGACCATTTCTTGTCAAGATATCAACACACCTCTCAAACTAGCACTGTACAGGT 775  
 21 AAlaAlaValaGlnGluLeuAsnProGluArgProGluYAlaGluGluArgVal 40  
 |||||  
 DB 776 GCTGCTGTCTCAGAGAGAGTGAACCTTGAGCGCCGCCAGCGCGGAGAGCGGGT 835  
 41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyLysArgProGly 60  
 |||||  
 DB 836 CCGGAGAGGACAGTAGTGCGCAATCGAGAGCGTCTCCCACTGGGTGGTCCGCGGG 895  
 61 ProGluGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnIlaGlyPro 80  
 |||||  
 DB 896 CCGGAGGAGGAGGAGGAGCTGGAATCCCAACCACTCTCCAGACCCAGCGCTGTCCA 955  
 81 GluSerSerCysLeuArgGlnGlyLysGlyGlnAsnGlyAspAspSerSerAlaGly 100  
 |||||  
 DB 956 GAATCTAGCTGCTGAGAGAGGCGGAGAGGCGCAAGATGGGAGAGACTGTCCTGCGC 1015  
 101 GlyAspPheProProProAlaGluValaGluProThrProGluAlaGluLeuAlaGln 120  
 |||||  
 DB 1016 GCGGACTTCCCGCCCGGACAGAGTGGCAACCGACCGCCGAGCGCGAGCTGCTCCGCG 1075  
 121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGlnGlu 140  
 |||||  
 DB 1076 CCTTGTCATGACTCCGAGGCGCGTAGTTGGGGGCTCTCGCCGAGGGGCGCAAGAGGAG 1135  
 141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysHisArgArgArgProSerLysLys 160  
 |||||  
 DB 1136 TGGGAGACAGACAGACAGACCTGGGGAAGAAAAACATAGAGAGCGCCCGTCAAGAA 1195  
 161 LysArgHisTrpLysProTrpTrpLysLeuThrTrpGlnGluLysLysLysPheAspGlu 180  
 |||||  
 DB 1196 AAGCGGCACTGGAAACCGCTACTACAGCTGAGCTGGGAAGAAAGAAAGTTCCAGCAG 1255  
 181 LysGlnSerLeuArgAlaSerArgLysArgAlaGlnMetPheAlaLysGlyGlnProVal 200  
 |||||  
 DB 1256 AAACAGAGGCTTCGAGCTTCAGAGATCCGAGCGGAGATGTTGCCAAGGCGCAAGCGGTC 1315  
 201 AlaProTyrAsnThrThrGlnPheLeuMetAspHisAspGlnGluProAspLeu 220  
 |||||  
 DB 1316 GCGCCCTATACACACCCAGTCTCTCATGTGATGATCAAGACAGAGAGCGGATCTC 1375  
 221 LysThrGlyLeuTyrSerLysArgAlaAlaLysSerAspAspThrSerAspAsp 240  
 |||||  
 DB 1376 AAACCGGCGCTGTACTCCAGCGGGCGCCCGCAATCCGAGACACACGATACAGAC 1435  
 241 PheMetGluGluGlyGlyGluAspGlyLysSerAspGlyMetGlyGlyAspGlySer 260  
 |||||  
 DB 1436 TTCATGAGAAAGAGGGGAGAGATGGGCGGACCATGGGATGGAGGGAGCGGACGCC 1495  
 261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280  
 |||||  
 DB 1496 GAGTTTCTGACAGCGGAGCTTCTCGGAGACGTACAGAGCGGTACACACGAGAGCTGCGAG 1555  
 281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluLeuGluLysCysLeuSerArg 300  
 |||||  
 DB 1556 AACATGAGCAAGCAGAGCTCATCAAGAGATCACTGGAACGTGGAAGATGCTCTCGCGCC 1615  
 301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAla 320  
 |||||  
 DB 1616 ATGAGGAGCAGAACACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGAGCAGACCGG 1675  
 321 ArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340  
 |||||  
 DB 1676 CGTGTGCGGAGAGCTGAGCTGAGCTGAGACCGGCTGCGCGGAGAACCTCCAGCTGTGTG 1735

QY 341 ThrGluAsnGluLeuHisArgGlnGluArgAlaProLeuSerLysPheGlyAsp 359  
 |||||  
 DB 1736 ACCGAGAACGAGACTGACCGGAGCAGAGGCGCGCCCTCCAGATTGGAGAC 1792

RESULT 5  
 AC024047/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

On Jun 16, 2000 this sequence version replaced gi:7109599.

Center: Washington University Genome Sequencing Center  
 Web site: http://genome.wustl.edu/gsc/index.shtml

Project Information  
 Center project name: H.NH0403G03  
 Summary Statistics

Sequencing vector: MJ3; 100%  
 Sequencing method: Plasmid; 0%  
 Chemistry: Dye-terminator Big Dye; 0% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 127942 bases at least Q40  
 Consensus quality: 133308 bases at least Q20  
 Insert size: 147000; agarose-1p  
 Insert size: 139626; sum-of-contigs  
 Quality coverage: 4.06 in Q20 bases; agarose-1p  
 Quality coverage: 4.33 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 28 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1575: contig of 1575 bp in length  
 1576 1675: gap of unknown length  
 1676 3114: contig of 1439 bp in length  
 3115 3214: gap of unknown length  
 3215 5389: contig of 2175 bp in length  
 5390 5489: gap of unknown length  
 5490 7129: contig of 1640 bp in length  
 7130 7229: gap of unknown length  
 7230 9240: contig of 2011 bp in length  
 9241 9340: gap of unknown length  
 9341 11177: contig of 1837 bp in length  
 11178 11277: gap of unknown length  
 11278 13113: contig of 1836 bp in length  
 13114 13213: gap of unknown length  
 13214 16526: contig of 3313 bp in length  
 16527 16626: gap of unknown length  
 16627 19864: contig of 3238 bp in length



PUBMED 12119119  
2 (bases 1 to 3402)  
REFERENCE Huang, F., Wagner, M. and Siddiqui, M.  
AUTHORS  
TITLE Submitted (19-Mar-2002) Anatomy and Cell Biology, State University  
JOURNAL of New York Health Science Center, 450 Clarkson Avenue, Brooklyn,  
NY 11203, USA

## FEATURES

## source

## gene

## CDS

Location/Qualifiers  
1..3402  
/organism="Mus musculus"  
/strain="129/SvJ"  
/db\_xref="taxon:10090"  
1..3402  
/gene="Clp1"  
566..1666  
/note="nuclear protein; CLP-1"  
/product="cardiac lineage protein 1"  
/protein\_id="AA09026.1"  
/db\_xref="GI:20135641"  
/translation="MAEPLLEHQPOTSNGTGAAYVHEHTSERPPSAEERPKED  
SRMGRASLQSGSPGEGEGGLKLPLOTNACPELSLEKGGKONGDLSTGA  
SPSAEGEMSESLVOPGHDSEATROEAPAGGEPWGOOROLGKKRRRPSKKRR  
WKPYKLTMEKKKKFDEKQSLASRVRAMAKOPVAPYNTOTFMDHNOEERDLK  
TGLYPRRAAKSDTSDDEYVEKEGEDGSDGMDGDSERLQRPSEYERYHAESL  
QMSNOELIKEYLEKLSRKEDENNRLKRLGSDARVRELELELDRLAENL  
QLTLENELHROERAPLSKFGD"

BASE COUNT 920 a 771 c 890 g 821 t  
ORIGIN

## Alignment Scores:

Pred. No.: 1.19e-75 Length: 3402  
Score: 1587.50 Matches: 308  
Percent Similarity: 88.86% Conservative: 11  
Best Local Similarity: 85.79% Mismatches: 37  
Query Match: 83.12% Indels: 3  
DB: 10 Gaps: 2

US-09-972-758a-2 (1-359) x AY090614 (1-3402)

QY 1 MetAlaGluProPheLeuSerGluTyrGlnHsGlnProGlnThrSerAsnGlyThrGly 20  
DB 596 ATGCCCGGACCACTTGTACAGAACATCAACCCGCTCAACATGACACTGTACAGT 655  
QY 21 AlaAlaValAlGlnGluLeuAsnProGluArgProGluAlaGluGluVal 40  
DB 656 GCTGCTGTTCATGAGACATACCTCTGAGCGCCCCCAAGCGCGAGAGAACGGGTG 715  
QY 41 ProGluGluAspSerArgTyrGlnSerArgAlaPheProGlnLeuGlyArgProGly 60  
DB 716 CCCAAGGAGACAGTGGTGGCAATCGAGACCGTCTTGCAGTCCGCTAGCCGTCCACGG 775  
QY 61 ProGluGluGluSerLeuGlnSerGlnProProPheLeuGlnThrGlnAlaCysPro 80  
DB 776 CAGAGGAGAGAGGGGCTTCAAGACACAGCTGCCCATTCGACAGCAATGCTGTCCA 835  
QY 81 GluSerSerCysLeuArgGluGlyGluGlyGlnAsnGlyAspAspSerSerAlaGly 100  
DB 836 GAATTGAGCTCCCTGAGAGAGGCGAGAAAGGCCAGAAATGGGAGAGACTTATCCACTGC 895  
QY 101 GlyAspPheProProPheAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120  
DB 896 GGT---GCCTCCCGTCCGCGAGAGAGCCGATGTCAGAGATCC-----CTCGTGCAG 946  
QY 121 ProCysHisAspSerGlnAlaSerLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140  
DB 947 CCAGGTCTATGACTGGAGGCGCACCAAGAGAGAGAGGCTCTGCGCGGAGGAGAGAGCCA 1006  
QY 141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160  
DB 1007 TGGGGAGACGCAACAGAGACACCTGTGGCGAAGAAAAACATCGAGAGAGCGCCCTCAAGAGAG 1066

QY 161 LysArgHisTrpLysProTyrTrpLysLeuThrTrpGluGluLysLysLysPheAspGlu 180  
DB 1067 AAGCGCATTTGGAGACCCCTACTACAGCTGCTGGAGAGAGAAAAAAGTTGACAGCAG 1126  
QY 181 LysGlnSerLeuAlaGlnAlaSerArgTyrLeuArgAlaGluMetPheAlaLysGlyGlnProVal 200  
DB 1127 AAGCAGAGCCCTGCGAGCTTCCGGGTTGACCCCGAGATGTTCCCAAGGGCCACCCAGTT 1186  
QY 201 AlaProTyrAsnThrThrGlnPheLeuMetLaspAspHisAspGlnGluGluProAspLeu 220  
DB 1187 GCGCCCTATTAACACACACCCACTTCTCTGATGATGACACGATCAAGAGAGACCTGATCTC 1246  
QY 221 LysThrGlyLeuTyrSerLysArgAlaAlaLysSerAspAspThrSerAspAspAsp 240  
DB 1247 AAAACCGGCTTTATACCCCAAGCGGACCCGCCCAATCCGACGACACACGAGATGAGAT 1306  
QY 241 PheMetGluGluGluGlyGluGluAspGlyLysSerAspGlyMetGlyLysGlySer 260  
DB 1307 TTTGTGGAGAGAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1366  
QY 261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280  
DB 1367 GAGTTTCTGACGCGGACTTCTCGGAGACGTACAGCGGCTACCAACGCGAGAGCTGTGAG 1426  
QY 281 AsnMetSerLysGlnGluLeuLeuLysGluTyrLeuGluGluGlyLysCysLeuSerArg 300  
DB 1427 AACATGACCAAGCAGAGACCTCATCAACAGTACCTGGAGCTGGAGAGAGTGCCTCCCGC 1486  
QY 301 MetGluAspGluAsnAsnArgLeuArgLeuGlnSerLysArgLeuGlyLysAspAspAla 320  
DB 1487 AAGCAGAGCAAAATACCGGCTGCGGTGAGAAAGCAAGCGGCTGGAGGCGTGTGACGCG 1546  
QY 321 ArgValArgGluLeuGluGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGluLeu 340  
DB 1547 CGAGTGGCGGAGCTGTGACCTGTGAGCTGAGCGGCTGCGGCTGGAACCTCAGCTGTG 1606  
QY 341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359  
DB 1607 ACCGAGAACGAATGACCGGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1663

## RESULT 7

AL731805 111002 bp DNA linear ROD 21-JUN-2002  
LOCUS  
DEFINITION Mouse DNA sequence from clone Rp23-358E19 on chromosome 11,  
complete sequence.

ACCESSION AL731805  
VERSION AL731805.8 GI:21540125  
KEYWORDS HMG.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 111002)  
AUTHORS Dunn, M.  
TITLE Direct Submission  
JOURNAL Submitted (21-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquerry@sanger.ac.uk; clone requests: clonerequests@sanger.ac.uk

COMMENT On Jun 23, 2002 this sequence version replaced gi:21531483.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquerry@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate



chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; SW, SwissProt; Tr, TREMBL; Wp, WormPEP; Information on the WormPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-358E19 is from the RPc1-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

FEATURES  
source  
1. 111002  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="11"  
/clone="RP23-358E19"  
/clone.lib="RPc1-23"  
VECTOR: PBACe3.6.

BASE COUNT 28051 a 27579 c 27135 g 28237 t  
ORIGIN

Alignment Scores:  
Pred. No.: 3.95e-74 Length: 111002  
Score: 1587.50 Matches: 308  
Percent Similarity: 88.86% Conservative: 11  
Best Local Similarity: 85.79% Mismatches: 37  
Query Match: 83.12% Indels: 3  
DB: 10 Gaps: 2  
US-09-972-758a-2 (1-359) x AL731805 (1-111002)

```
QY      1 MetLaGluPrPheLeuSerGluTrGlnHisGlnProGlnThrSerAsnCysThrGly 20
DB      91440 ATGGCCGACGACCTTGACGACCAATCACCACGAGCTTCAACTAGCACTGACAGT 91499
QY      21 AlaAlaIaValGlnGluGluLeuAsnProGluArpProGlyAlaGluGluArGVal 40
DB      91500 GCTGCTGTGTCCTGAAGACATACCTCTGAGCGCCCCAACAGCGGAGCAAGGGGTG 91559
QY      41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyArgProGly 60
DB      91560 CCCAAGAGGAGCAGTAGGTGGCATTGACGACCTGCTTGCAGTCCGGTAGCCGTCAGGG 91619
QY      61 ProGluGlyGluGlySerLeuGlnSerGlnProProLeuGlnThrGlnAlaCysPro 80
DB      91620 CAGAGGAGAGAAAGGGGCTTGAACACAGCTGCTCCCATTTGACAGCAATGCTGTCCA 91679
QY      81 GluSerSerCysLeuArGluGlnGlyGlnGlyGlnAsnGlyAspAspSerSerAlaGly 100
DB      91680 GAATTGAGCTTCCTGGAGAAAGGGGAGAGAGGCGCAGAAATGGGAGACTTATCCACTGGC 91739
QY      101 GlyAspPheProProAlaGluValGlnProThrProGluAlaGluLeuLeuAlaGln 120
DB      91740 GGT---GCCGCCCGCTGGCGGAGAGAGCGCATGTCAGAGTCC-----CTGCTGCAG 91790
QY      121 ProCysHisAspSerGlnAlaSerLysLeuGlyAlaPheAlaAlaGlyGlyGluGlu 140
DB      91791 CCAGGTATGACTCTGGAGGCGCACCAAGAGAGGCTCTCTCCCGCGGAGAGCGAGCAACA 91850
QY      141 TrpGlyGlnGlnGlnArGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
DB      91851 TGGGACACGACACAGACACACTGGGCAAGAAAAACATCGGACACGCCCTTCAAGAGAG 91910
QY      161 LysArgHisTrpLysProTyrTrpLysLeuThrTrpGluGluLysLysLysPheAspGlu 180
DB      91911 AAGCGCATTTGGAGCCCTACTACAGCTGACTTGGAGGAGAGAAAAAATTCGACGAG 91970
QY      181 LysLysInsLeuArGAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
DB      91971 AAGCAGAGCTTGAGAGCTTCCGGGTTGACGCCGAGATGTTCCCAAGGGCGCACGAGTT 92030
```

```
QY      201 AlaProTyrAsnThrTrpGlnPheLeuMetAspAspHisAspGlnGluProAspLeu 220
DB      92031 GCGCCCTTATACACACAGCAGCTTCTCATGATGATACACAGTCAGAGAGACCTGATCTC 92090
QY      221 LysThrGlyLeuTyrSerLysArgAlaAlaLysSerAspAspThrSerAspAsp 240
DB      92091 AAAACGGGCTTTATCCCAAGCGGAGCGGCCCAATTCGACGACACCGAGATGAGAT 92150
QY      241 PheMetGluGluGlyGlyGluGluGluAspGlyLysSerAspGlyMetGlyLysArg 260
DB      92151 TTTCTGGAAGAAAGCTGTGAGAGAGACGAGCGCATGTGATGAGAGCGGAGCGCAGC 92210
QY      261 GluPheLeuGlnArGAspPheSerGluThrTrpGluArgTrpHisThrGlnSerLeuGln 280
DB      92211 GAGTTTGTGACGGCGGAGTTCTCGAGACGTAACAGCGGTACACGCCGAGAGCTTCGAG 92270
QY      281 AsnMetSerLysGlnGluLeuIleLysGluTyrLeuGluGluGlyLysCysLeuSerArg 300
DB      92271 AACATGAGCAAGCAGAGAGCTCATCAAGAGTACCTGGAGCTGGAGAAAGTGTCTCCCGC 92330
QY      301 MetGluAspGluAsnAsnArgLeuArgLeuGlnSerLysArgLeuGlyLysAspAla 320
DB      92331 AACGAGAGCAGAAATAACCGCTCGGCTGAAAGCAAGCGGCTGGAGCGCTGACGCG 92390
QY      321 ArgValArgGluLeuGluGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeu 340
DB      92391 CGAGTGGGGGAGCTCTGAGCTGAGCTGAGCGGCTGGCGCTGAGAACCTTCACACTCTGTG 92450
QY      341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359
DB      92451 ACCGAGAACGAACTGCACCGGCGACGAGGAGCGCGCTTTTCCAAAGTTGGCGGAC 92507
```

RESULT 8  
LOCUS BC022111 1488 bp mRNA linear ROD 07-AUG-2002  
DEFINITION Mus musculus, similar to HMBA-inducible, clone MGC:36245  
ACCESSION BC022111 IMAGE:5067620, mRNA, complete cds.  
VERSION BC022111.1 GI:18381039  
KEYWORDS MGC.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
REFERENCE 1 (bases 1 to 1488)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (24-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
[info@bgsc.bc.ca](mailto:info@bgsc.bc.ca)

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Buterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Leticia Hsiao, Martin Kirzylinski, Reta Kusche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saedi, Jacqueline Schein, Duane Smalins, Michael Smith, Lorraine Spence, Jeff Sloty, Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 57 Row: a Column: 15

This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency OK analysis, similarity but not identity to protein.

## FEATURES

Location/Qualifiers

1..1488  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /map="CECH.11"  
 /clone="MGC:36249 IMAGE:5067620"  
 /tissue\_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."  
 /clone\_1ib="NCI CGAP\_Lu29"  
 /lab\_host="DH10B"  
 /note="vector: pcwv-sport6"  
 /codon\_start=1  
 /product="Similar to HMB-Inducible"  
 /protein\_id="AAH2211.1"  
 /db\_xref="GI:18381040"  
 /translation="MSESLVDPHDSSEATKQEPAPAGEEPEWQOQOLGKKHRRP  
 SKKHHMPYKLTWEKKKFEDEKSLASRVRAEMKGGPVAPNTTQFIMDDHQ  
 EEPDLKTLGPRRAAKSDTSDDEFEAGEEDGSDMGDSSEFLOPFSFYER  
 YHBSLQNMKSQOELKEYLELEKCSRREDENNRLRLBSKRLGVDARVRELELDR  
 LRAENIOLITENEHLRQOEAPLRSKRGD"

## CDS

BASE COUNT 385 a 371 c 449 g 283 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 1.75e-75 Length: 1488  
 Score: 1577.50 Matches: 305  
 Percent Similarity: 88.86% Conservative: 14  
 Best Local Similarity: 84.96% Mismatches: 37  
 Query Match: 82.59% Indels: 3  
 DB: 10 Gaps: 2

US-09-972-758A-2 (1-359) x BC022111 (1-1488)

OY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnGlyThrGly 20  
 DB GTCTCACAGCCACTTCTGACAGAACATCACACCGCTCAAACTAGACACTGTACAGT 177  
 OY 21 AlaAlaValaGlnGluLeuAsnProGluArgProGluAlaGlnGluVala 40  
 DB GCTGCTGTTTCATGAAGACATACCTCTGACGCCCCCAAGCGCGAGAGAGGGTG 237  
 OY 41 ProGluGluAspSerArgTyrGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60  
 DB CCCAAGGAGGACAGTAGTGGCAATCGAGAGCGTCTTCCGCGTAGCGTCCAGCG 297  
 OY 61 ProGluGlnGlySerLeuGlnSerGlnProProProLeuGlnThrGlnAlaGlyPro 80  
 DB CAGAGAGGAGAGGGCGGCTGAAACACACAGCTGCCCATTCGACAGCAATGCTGTCCA 357  
 OY 81 GluSerSerCysLeuArgGlnGluGlyGlnAsnGlyAspAspSerSerAlaGly 100  
 DB GAATTTAGGCTCCCTGGAGAGAGCGGAGAGGCGCAATGGGAGACTTATCCACTGGC 417  
 OY 101 GlyAspPheProProProAlaGlnValaGlnProThrProGluAlaGlnLeuAlaGln 120  
 DB GGT---GCCCTCCCGCTGGCGGAGAGAGCGCATGTACAGATCC-----CTCGTGCAG 468  
 OY 121 ProCysHisAspSerGlnAlaSerLeuGlnGlyAlaProAlaAlaGlyGlnGluGlu 140  
 DB CCAAGTATGACTCTGGAGCGCCCAACAGAGCGCTCTGCCCGGAGAGGAGAGAGCA 528  
 OY 141 TrpGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 160  
 DB TGGGACAGACAGACAGACAGCTGGCGCAAGAAAAAATCGAGAGAGCCCTCAAGAG 588  
 OY 161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGlnGluGlnGlnGlnGlnGln 180  
 DB AAGCGCATTTGGAAGCCCTACTACAGCTGAGTGGAGAGAGAGAGAGAGAGAGAG 648

OY 181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlnProVal 200  
 DB AAGCAGAGCCCTGGAGCTTGGCGGTTGACGCCGAGATTTGCCAAGGCCAGCCAGT 708  
 OY 201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluProAspLeu 220  
 DB GCGCCCTATTAACACACAGCAGTTCTCATGATGACACAGATCAGAGAGAGCTTATCTC 768  
 OY 221 LysThrGlyLeuTyrSerTyrArgAlaAlaAlaLysSerAspAspThrSerAspAsp 240  
 DB AAAACCGGCTTTTACCAGGCGGAGCGGCCCAATCCGACACACAGCATGAGAT 828  
 OY 241 PheMetGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260  
 DB TTTGTGAAGAGCTGTGTAGACAGACGAGCAGCATGCGATGGAGGAGGAGCGCAGC 888  
 OY 261 GluPheLeuGlnArgAspPheSerGluThrTyrGluAlaGlyTyrHisThrGlnSerLeuGln 280  
 DB GAGTTTCTGCACCGGCTTCTTCGACAGCTACGAGCGGTACACCGCAGAGCTCCAG 948  
 OY 281 AsnMetSerLysGlnGlnLeuLeuLysGluTyrGlnGlnGlnGlnGlnGlnGlnGlnGln 300  
 DB AACATGACCAACAGCAGAGCTCTTCAAGAGTACTGTGAGCTGAGAGAGTGCCTTCCCGC 1008  
 OY 301 MetGluAspGluAsnAsnArgLeuArgLeuGlnSerLysArgLeuGlyGlyAspAspAla 320  
 DB AAGGAGAGCAGAAATTAACCGGCTGCGTGGAACACAGCGCTGGAGGCTGCACCG 1068  
 OY 321 ArgValaArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 340  
 DB CGAGTGGCGGAGCTTGAGCTGAGCTGCAGCGGCTGCGCTGAGAACCTCAGCTGCTG 1128  
 OY 341 ThrGluAsnGlnLeuHisArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 359  
 DB ACCGAGAGACGACGTGACCGGCGAGGAGCGAGCGGCTCTTTCACAGTTGCGGAGC 1185

## RESULT 9

AC120950 175120 bp DNA linear HTG 23-JUN-2002  
 AC120950 LOCUS  
 Rattus norvegicus clone CH230-456H21. \*\*\* SEQUENCING IN PROGRESS

AC120950.2 GI:21902729  
 VERSION  
 HTG: HTGS\_PHASE1.  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Rattus norvegicus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

## REFERENCE 1 (bases 1 to 175120)

1. Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alsbrooks, S.L., Amaralunge, H.C., Are, J.R., Ayale, M., Banks, T., Barbara, J., Benton, J., Bimege, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Briteva, M., Brown, E., Brown, M., Bryant, N.P., Bunay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Cartron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dabholme, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homsl, F., Howard, S., Huber, J., Huiyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichteage, O., Lien, C., Liu, C., Liu, J., Liu, W., Louieged, H.,

	REFERENCE AUTHORS	REFERENCE TITLE	REFERENCE JOURNAL
*	Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, L., Martindale, A., Martinez, E., Massey, E., McWhirley, E., McLeod, M.P., Meador, M., Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J.P., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nookankwong, S., Ogih, M., Okunubu, G., Ordungwe, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojudoacan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshbari, N., Sisson, J., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Tanton, A., Vatek, A., Tabar, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telirod, B., Thomas, R., Thomas, S., Usmani, K., Vasquez, L., Vera, Y., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.	Direct Submission Unpublished 2 (bases 1 to 175120)	
	Worley, K.C.	Submitted (14-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 3 (bases 1 to 175120)	
	Worley, K.C.	Direct Submission Submitted (23-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA On Jul 18, 2002 this sequence version replaced gi:20564294.	
COMMENT		----- Genome Center ----- Center: Baylor College of Medicine Center code: BCM Web Site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc-help@bcm.tmc.edu Project Information Center project name: GWRO Center clone name: CH230-456H21 ----- Summary Statistics ----- Sequencing vector: Plasmid; Chemistry: Dye-terminator Big Dye; 100% of reads Assembly program: Phrap; Version 0.990329 Consensus quality: 147564 bases at least Q40 Consensus quality: 147590 bases at least Q30 Consensus quality: 152266 bases at least Q20 ----- NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html) NOTE: This is a "working draft" sequence. It currently consists of 41 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.	
*	1	1114: contig of 1114 bp in length	
*	1115	1214: gap of unknown length	
*	1215	2688: contig of 1474 bp in length	
*	2689	2788: gap of unknown length	
*	2789	4582: contig of 1794 bp in length	
*	4583	4682: gap of unknown length	
*	4683	6524: contig of 1842 bp in length	
*	6525	6624: gap of unknown length	
*	6625	8342: contig of 1718 bp in length	
*	8343	8442: gap of unknown length	
*	8443	11551: contig of 3109 bp in length	
*	11552	11651: gap of unknown length	
*	11652	13499: contig of 1848 bp in length	
*	13500	13599: gap of unknown length	

	*	13600	14781..contig of 1182 bp in length
	*	14782	14881..gap of unknown length
	*	14882	17066..contig of 2185 bp in length
	*	17067	17166..gap of unknown length
	*	17167	19189..contig of 2023 bp in length
	*	19190	19289..gap of unknown length
	*	19290	20786..contig of 1497 bp in length
	*	20787	20886..gap of unknown length
	*	20887	22783..contig of 1897 bp in length
	*	22784	22883..gap of unknown length
	*	22884	24379..contig of 1496 bp in length
	*	24380	24479..gap of unknown length
	*	24480	25904..contig of 1425 bp in length
	*	25905	26004..gap of unknown length
	*	26005	28572..contig of 2568 bp in length
	*	28573	28672..gap of unknown length
	*	28673	31006..contig of 2334 bp in length
	*	31007	31106..gap of unknown length
	*	31107	33737..contig of 2631 bp in length
	*	33738	33837..gap of unknown length
	*	33838	35985..contig of 2148 bp in length
	*	35986	36085..gap of unknown length
	*	36086	39468..contig of 3383 bp in length
	*	39469	39568..gap of unknown length
	*	39569	43008..contig of 3440 bp in length
	*	43009	43108..gap of unknown length
	*	43109	45505..contig of 2397 bp in length
	*	45506	45605..gap of unknown length
	*	45606	48876..contig of 3271 bp in length
	*	48877	48976..gap of unknown length
	*	48977	52243..contig of 3267 bp in length
	*	52244	52343..gap of unknown length
	*	52344	56160..contig of 3817 bp in length
	*	56161	56260..gap of unknown length
	*	56261	60475..contig of 4215 bp in length
	*	60476	60575..gap of unknown length
	*	60576	64722..contig of 4147 bp in length
	*	64723	64822..gap of unknown length
	*	64823	68616..contig of 3794 bp in length
	*	68617	68716..gap of unknown length
	*	68717	73739..contig of 5023 bp in length
	*	73740	73839..gap of unknown length
	*	73840	75346..contig of 5507 bp in length
	*	75347	79446..gap of unknown length
	*	79447	83302..contig of 3856 bp in length
	*	83303	83402..gap of unknown length
	*	83403	90225..contig of 6823 bp in length
	*	90226	90325..gap of unknown length
	*	90326	96171..contig of 5846 bp in length
	*	96172	96271..gap of unknown length
	*	96272	101663..contig of 5392 bp in length
	*	101664	101763..gap of unknown length
	*	101764	108594..contig of 6831 bp in length
	*	108595	108694..gap of unknown length
	*	108695	114283..contig of 3589 bp in length
	*	114284	114383..gap of unknown length
	*	114384	120900..contig of 6517 bp in length
	*	120901	121000..gap of unknown length
	*	121001	129340..contig of 8340 bp in length
	*	129341	129440..gap of unknown length
	*	129441	138422..contig of 8882 bp in length
	*	138423	138522..gap of unknown length
	*	138523	149684..contig of 11162 bp in length
	*	149685	149784..gap of unknown length
	*	149785	160420..contig of 10636 bp in length
	*	160421	160520..gap of unknown length
	*	160521	175120..contig of 14600 bp in length.
FEATURES			
Source			Location/DnaIIFields 1..175120 /organism="Rattus norvegicus" /db_xref="taxon:10116" /clone="C41556.g_43211 t_6647 others"
BASE COUNT			41850 a 41856 c41556 g 43211 t 6647 others

## ORIGIN

## Alignment Scores:

Pred. No.: 4,59e-70 Length: 175120  
 Score: 1514.50 Matches: 298  
 Percent Similarity: 86.91% Conservative: 14  
 Best Local Similarity: 83.01% Mismatches: 44  
 Query Match: 79.29% Indels: 4  
 DB: 2 Gaps: 2

US-09-972-758a-2 (1-359) x AC120950 (1-175120)

```

OY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAncySthrgly 20
DB 72106 ATGGCTGAGCCACTCTTGTCAACATCAACACGCTCAACATCAACATCAACATCAACAT
OY 21 AlaAlaValAlaGlnGluLeuAsnProGluAraGProProGlyValAlaGluGlnAraVal 40
DB 72166 GCTGCTGTTCATGAGAGAGAGAGAACTGTGAGCCGCCCAAGCGCGGAGAGAGAGGCTG 72225
OY 41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyAraProGly 60
DB 72236 CCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 72285
OY 61 ProGluGluGlySerLeuGlnSerGlnProProProLeuGlnThrGlnAlaCysPro 80
DB 72286 CAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 72345
OY 81 GluSerSerCysLeuArgGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 100
DB 72346 GAATTGAGCTGCTGTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 72405
OY 101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
DB 72406 GGT---GCCTCCCTTCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 72456
OY 121 ProCysHisAspSerGluAlaSerLysLeuGlnAlaProAlaAlaGlyGlyGluGluGlu 140
DB 72457 CCAGGTCACTGACTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 72516
OY 141 TrpGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 160
DB 72517 TGGGAGACAGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 72576
OY 161 LysArgHisTrpLysProTyrTrpLysLeuThrTrpGluGluGluGluGluGluGluGluGlu 180
DB 72577 AAGGACATTTGAAACCTTATTACAAAGGTGAC-TGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 72635
OY 181 LysGlnSerLeuArgAlaSerArgLysArgLysArgLysArgLysArgLysArgLysArgLys 200
DB 72636 AAGCAAAAGCCGGGAGCTTCCGCAATTCGAGCCGAGATGTTCCCAAGAGGCGCAGCGGTT 72695
OY 201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220
DB 72696 GCGCGGTATTAACACACAGCACTTCCTCATGATGATCATGATCATGATCATGATCATGATCAT 72755
OY 221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAsp 240
DB 72756 AAAACCGGCTTTTACCTTAACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 72815
OY 241 PheMetGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 260
DB 72816 TTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 72875
OY 261 GluPheLeuGlnAraGspPheSerGluThrTyrGluAraGlyTrpHisThrGlnSerLeuGln 280
DB 72876 GAGTTTCTGACGCGGAGACTTTCCGAGACTTACGAGCGGTACCAAGCGCGAGAGAGAGAGAG 72935
OY 281 AsnMetSerLysGlnGluLeuLeuLysGluTyrLeuGlnGluGluGluGluGluGluGluGlu 300
DB 72936 AACATGAGCAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 72995
OY 301 MetGluAspGluAsnAsnArgLeuArgLeuGlnSerLysArgLeuGlyGlyAspAspAla 320
  
```

```

DB 72996 AAGGAGAGCAAAATTAACCGGCTGCGGTGGAAGAGAGAGAGAGAGAGAGAGAGAGAG
OY 321 ArgValArgGluLeuGlnGluLeuLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340
DB 73056 CAGATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
OY 341 ThrGluAsnGluLeuHisArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 359
DB 73116 ACCGAAAGCACTGACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
  
```

RESULT 10  
 AX331371 461 bp DNA Linear PAT 09-JAN-2002  
 LOCUS AX331371  
 DEFINITION Sequence 1880 from Patent WO0194629.  
 ACCESSION AX331371  
 VERSION AX331371.1 GI:18122005  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1  
 AUTHORS Young, P. E., Augustus, M., Carter, K. C., Ehner, R., Endress, G.,  
 Horrigan, S., Soppet, D. R. and Weaver, Z.  
 TITLE Cancer gene determination and therapeutic screening using signature  
 gene sets  
 JOURNAL Patent: WO 0194629-A 1880 13-DEC-2001;  
 Avalon Pharmaceuticals (US)  
 FEATURES  
 source Location/Qualifiers  
 1..461  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 BASE COUNT 105 a 134 c 138 g 80 t 4 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 9.07e-26 Length: 461  
 Score: 629.00 Matches: 120  
 Percent Similarity: 93.85% Conservative: 8  
 Best Local Similarity: 92.31% Mismatches: 2  
 Query Match: 32.93% Indels: 1  
 DB: 6 Gaps: 0

US-09-972-758a-2 (1-359) x AX331371 (1-461)

```

OY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAncySthrgly 20
DB 72 ATGGCCGAGCCATTCTTCTCAGATATCAACACAGCTCAAACTGACACTGTACAGT 131
OY 21 AlaAlaValAlaGlnGluLeuAsnProGluAraGProProGlyValAlaGluGlnAraVal 40
DB 132 GCTGCTGTTCACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 191
OY 41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyAraProGly 60
DB 192 CCCGAGGAGCAGTAGCTGCAATTCGAGAGAGCTTCCCGAGTGGGTGCGCCGCGG 251
OY 61 ProGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 80
DB 252 CCGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 311
OY 81 GluSerSerCysLeuArgGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 100
DB 312 GAATCTAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 371
OY 101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
DB 372 GC-GACTTCCCGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 130
OY 121 ProCysHisAspSerGluAlaSerLysLeu 130
DB 431 CTTGTATCATCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 460
  
```

RESULT 11  
AK056946 1330 bp mRNA linear PRI 01-AUG-2002  
LOCUS AK056946  
DEFINITION Homo sapiens cDNA FLJ23284 f1s, clone SKMUS1000104, weakly similar  
to Homo sapiens mRNA for HEXIM1 protein, complete cds.  
ACCESSION AK056946.1 GI:16552484  
VERSION AK056946.1  
KEYWORDS oligo capping; f1s (full insert sequence).  
SOURCE Homo sapiens skeletal muscle cDNA to mRNA, clone\_lib:SKMUS1  
clone:SKMUS1000104.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS 1  
Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K.,  
Arita,M., Musashino,K., Yuuki,H., Hara,H., Sugiyama,T., Irie,R.,  
Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,  
Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,  
Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,  
Magakura,M., Murakawa,K., Kanehori,K., Sugiyama,A., Kawakami,B.,  
Suzuki,Y., Sugano,S., Nagahari,K., Masuo,Y., Nagai,K. and  
Isogai,T.  
TITLE NEDO human cDNA sequencing project  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1330)  
AUTHORS Isogai,T., Otsuki,T. and Sugiyama,T.  
JOURNAL Direct Submission  
TITLE Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,  
Genomics Laboratory: 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail:genomics@h1.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan: cDNA full insert sequencing:  
Research Association for Biotechnology (RAB): cDNA library  
construction: Helix Research Institute (HRI) (supported by Japan  
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: RAB and  
HRI.  
FEATURES  
source  
1. location/qualifiers  
1..1330  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="SKMUS1000104"  
/tissue\_type="skeletal muscle"  
/clone\_lib="SKMUS1"  
/note="Cloning vector: pME18SfL3"  
240..1100  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BA071319.1"  
/db\_xref="GI:16552485"  
/translation="MMAITPNOTACNAESPALAEBAKTSQAPGSRQTPPEPHDSCSLP  
LTPRMESHSEDEDLGAVGGGLGNMRSPTQSPGSAEVLAKKRRRPSKRRRM  
PRLLEISWAKQDERQKQSRASRVRLEMAKQPAVYNTQFLNMDRDEPNLDV  
PHQISHGSGEAGEGSDGRAGHGFQKDFSEYERHTESLQGRSKQELVRLD  
ELKRLSQAEEETRLQOLQACTGQSCROVEELAAVQLRTENQRLQENQWMNE  
GCRDDEPPT"  
BASE COUNT 295 a 370 c 452 g 213 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1.82e-21 Length: 1330  
Score: 556.50 Matches: 141  
Percent Similarity: 51.40% Conservative: 42  
Best Local Similarity: 39.61% Mismatches: 118  
Query Match: 29.14% Indels: 55  
DB: Gaps: 9  
US-09-972-758a-2 (1-359) x AK056946 (1-1330)  
OY 20 G|VAlAAlAAlAValGInGInGInLeuAsnProGInuArProProGInAlaGInuGInuArG 39  
||||||| :|||

Db 87 GGAGCAGACAGAGAMAGCCGCCCTGATCCATTTGGCCCTTGGCTGGAGCTGT 146  
OY 40 ValProGInuGInuAspSerArgTTPGInSerArgAlaPheProGInuGInuGInuArgPro 59  
Db 147 -----GAAACAGCGGTGAGGACCGCTTTCGGGGCTGCATT----- 185  
OY 60 G|YProGInuGInuGInuSerLeuGInuSerGInu----- 70  
Db 186 -----GAGATTAAGGCTGCTACTAGTTCAGAGGCGTCTGTAAGATTTGGAACAGAG 239  
OY 71 -----ProProProLeuGInuThrGInuAlaCysProGInuSerSerCysLeuArgGInu 88  
Db 240 ATGATGCGCACCTCCGAAACAGACCGCTTGAATGACAGTACAGTGGCTTGGAGAG 299  
OY 89 G|uLysG|YInuSngInuAspAspSer-----SerAlaGly 100  
Db 300 GCCAAGACCTCTGTGTCGCCCGGAGGAGCCCCAAACACCCCTGAGCTCATGACTCTGT 359  
OY 101 G|YAspPheProProProAlaGInuAlaGInuProThrProGInuAlaGInuLeuAlaGIn 120  
Db 360 GGTTCCTGCGCCCTGACACCGCGGATGAGAGCCACACAGAGATGAAGATCTTGTCTGG 419  
OY 121 ProCys-----HisAspSerGInuAlaSerLysLeuGInuAlaProAlaAlaGly 136  
Db 420 GCTGTGCGTGCGCTGGCTGGACAGTGAAGATCCCGGACCCAGACCGAGGCGCTGC 479  
OY 137 G|YGInuGInuGInuProGInuGInuGInuGInuGInuGInuGInuGInuGInuGInu 156  
Db 480 TCACCGGAG-----GCTGTGCTGGCGCCGGAAGAAACCGCTGGCGG 521  
OY 157 ProSerLysLysLysArgHisThrLysProTyrLysLeuThrTPGInuGInuLysLys 176  
Db 522 CCATCGAAGCGCAAAAGCGACTGGCGACCTTACTGAGAGTGAAGTGGCTGCTGCAACAA 581  
OY 177 LysPheAspGInuLysGInuSerLeuArgAlaSerArgLleArgAlaGInuMetPheAlaLys 196  
Db 582 CAGCGGATGAGAGGACGACGACGAGGCGCTCCCGGTCGCGAAGATGTTCCGCAAA 641  
OY 197 G|YGInuProValAlaProTyrArgSerThrGInuPheLeuMetAspAspHisAspGInu 216  
Db 642 GCCAGCGCGTGGCGCCCTACAAACACCAACCAAGTCTGTAAGTACAGAGACCGGAG 701  
OY 217 G|uPro-----AspLeuLysThrGlyLeuTyrSerLysArgAlaAlaLysSerAsp 234  
Db 702 GAGCCCAACTTGGATGTCGCCCATGCGATCTCCACCCAGCTTCCACTGGCGAGAGT--- 758  
OY 235 AspThrSerAspAspAspMetGInuGInuGInuGInuGInuGInuGInuGInuGInu 254  
Db 759 -----GAGGCGGCGGACAGTGAAGG 779  
OY 255 MetG|YLysAspGInuSerGInuPheLeuGInuArgAspPheSerGInuThrTyrGInuArgTyr 274  
Db 780 CGGGGCCGAGCCGACGCTGACTTCCACGGGAAGAGACTTCTGTGACTTACGAACGCTTC 839  
OY 275 HisThrLysLeuSerLeuGInuAsnMetSerLysGInuLeuLeuLysLysLysLysLys 294  
Db 840 CACACCGACAGACCGTCGACGGCGCGAGCAACAGAGAGCTGTGCGAGACTACCTCGAGCTG 899  
OY 295 G|uLysCysLeuSerArgMetGInuAspGInuAsnAsnArgLeuArgLeuGInuSerLysArg 314  
Db 900 GAGAACCGCGCTGTGCGAGCGGAGAGAGACTAGAGAGCTGCGAGCTGCGAGCGCTGC 959  
OY 315 LeuG|YGInuLysAspAlaArg-----ValArgGInuLeuGInuGInuLeuAspArgLeuArg 333  
Db 960 ACCGCGCAGCAGTCTGCTCCCGCAGGTGGAGAGCTGCTGCGAGGTCGACGAGCTTCGG 1019  
OY 334 AlaGInuAsnLeuGInuLeuThrGInuAsnGInuLeuHisArgGInuGInu 349  
Db 1020 ACCGAACACGCGCGCTTCTGTCAGAGACACAGATGTGGAACGAGAG 1057  
RESULT 12  
AC008105  
LOCUS AC008105 116035 bp DNA linear HTG 27-JUN-2002

DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
Homo sapiens chromosome 17 clone CTD-2020K17 map 17, *** SEQUENCEINCOMPLETE IN PROGRESS ***	AC008105	24	GI:21617753	HTG: HTGS_PHASE1; HTGS-FULLTOP; HTGS-ACTIVEFIN.	Homo sapiens human.
REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS
1 (bases 1 to 116035)	Birren,B., Nusbaum,C. and Lander,E.	Homo sapiens chromosome 17, clone CTD-2020K17	Unpublished	2 (bases 1 to 116035)	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckler,J.R., Benn,J., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collamore,A., Cooke,P., DeBellano,K., Depayot,E., Devon,K., Dewar,K., Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G., Haeos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L., Karats,A., Lechoczy,J., Lien,C., Locke,K., MacDonald,P., Marquis,N., McEwan,P., McCurt,A., McKernan,K., McLaughlin,J., Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Naylor,J., Nilloff,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P., Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A., Tesfaye,S., Torturella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission	Submitted (22-JUN-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	3 (bases 1 to 116035)	Birren,B., Nusbaum,C., Lander,E., Allen,N., Anderson,S., Barna,N., Baatien,V., Bloom,T., Boguslavskiy,L., Boukhalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A., Cook,A., Cooke,P., DeBellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Gird,S., Graham,L., Grand-Pierre,N., Haeos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A., Karats,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., MacDonald,P., Major,J., Matthews,C., McCarthy,M., Meldrum,J., Menus,L., Mihova,T., Mieng,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schuppacker,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zaidoun,J., Zembek,L., Zimmer,A. and Zody,M.		
Direct Submission	Submitted (27-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	On Jun 27, 2002 this sequence version replaced gi:20148201.	All repeats were identified using RepeatMasker:	Smith, A.F.A. & Green, P. (1996-1997)	http://ftp.genome.washington.edu/RM/RepeatMasker.html
Genome Center	Center: Whitehead Institute/ MIT Center for Genome Research	Center code: WIBR	Web site: http://www-seq.wi.mit.edu	Contact: sequence.submissions@genome.wi.mit.edu	Project Information
Center project name: LB890	Center Clone name: 2020_K_17				

\* NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence

[illegible]

```

Db      18224 -----GAGGCGGGGACACTGATGGCGGGGCCGACGCCA 182538
Oy      259 ySeRgluPheLleuGlaArgAsPheSeRgluThrYrGluArgYrHisThrGluSerLe 279
Db      18259 CGGGTAGTTCTCCACGCGAGCACTTCTCTGTAGACTTACGAACGTTTCCACACGAGAGCTT 18318
Oy      279 uGlnAmMetSerLysGlnGluLeuLeuLysGluYrTrfLeuGluLeuGluLysCysLeuSe 299
Db      18319 GCAGGCGCGGACGACAGACAGACTGTGTCGAGACTACTCTGCACTGTGAAGCTTGAAGAGCGCTGTC 18378
Oy      299 rArMeLelGuAsPgluAsnArnArgrLeuArgrLeuGluSerLysArgrLeuGlyGlyAsPas 319
Db      18379 GCAGGCGGAGAGGAGCACTAGAGGCTGCAGCGACTGTGCAGCGCTGCACCGCGCAGGACATC 18438
Oy      319 pAlaArg---ValArgGluLeuGluGluLeuLeuAsPArgrLeuArgrLeuAlaGluAsnLeuGl 338
Db      18439 CTCGCCGACAGTGTGAGAGAGCTGGCTGTCCGAGGTCGAGAGCTCCGACCCGAAACACGAGCG 18498
Oy      338 nLeuLeuThrGluAsnGluLeuHisArgrGlnIn 349
Db      18499 GCTTCGTCAGAGACACCACTGTGCAACCGAGAG 18532

RESULT 13
LOCUS   AL662804
DEFINITION Mouse DNA sequence from clone RP23-341C5 on chromosome 11, complete
          sequence.
ACCESSION AL662804
VERSION   AL662804.17 GI:21621620
KEYWORDS  HTG.
SOURCE    mouse mouse.
ORGANISM  Mus musculus
          Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 223442)
          Blakey,S.
          Direct Submission
          Submitted (26-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
          Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
          humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
          On Jun 27, 2002 this sequence version replaced gi:21615602.

COMMENT  ----- Genome Center
          Center: Wellcome Trust Sanger Institute
          Center code: SC
          Web site: http://www.sanger.ac.uk
          Contact: humquery@sanger.ac.uk
          -----

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw,
SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information from the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep
constructed by the group of Piter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBac3.6.
location/Qualifiers
1..223442
/organism="Mus musculus"

```

```

/db_xref="taxon:10090"
/clone="RP23-341C5"
/clone_id="RPC1-23"

BASE COUNT      56058 a 54948 c 56347 g 56089 t
ORIGIN

Alignment Scores:
Pred. No.:      6,95e-19      Length:      223442
Score:          550.00      Matches:      157
Percent Similarity: 44.84%      Conservative: 47
Best Local Similarity: 34.51%      Mismatches: 108
Query Match:      28.80%      Indels:      143
DB:              10      Gaps:      16

US-09-972-758A-2 (1-359) x AL662804 (1-223442)

OY      10      GlnHscgInProglInThSerAsnCysThrClYAlaIaIaValGInGluLeuAsn 25
          ||| ||||| ||||| |||
Db      3186      CAAGAGCCAGCCGCTGCTATGAGAGATGCTATGTCAAAAAAAAAAAAAAAAAAAAA 3245
OY      30      ProGluAArgProProGluYlaGluGluArgValProGluGluAspSerArg-----"rP 47
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      3246      AAGAGAAA-----AACAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 3296
OY      48      GlSerArgAlaPheProGluLeuGly-----Gly 57
          ::::: ||| ||||| |||||
Db      3297      CGCAGCGCTTTAAATCCCGACACTCGCGAGCGACAGCGAGTGGATTTCTAACTTTGAGGC 3356
OY      58      ArgProGluProGluGlyGlySerLeuGluSerGlnProProProLeuGlnThGln 77
          ::||| ||||| ||||| |||||
Db      3357      CAGCCTGGTCTACAAAGTAGTGCCAG-----GACAGCCAG 3392
OY      78      AlaCysProGluSerSerCysLeuArgGluGlyLys-----LysGlyGln--- 89
          ||| ||||| ||||| ::|
Db      3393      GGCTACACAGAGAAACCCTGTCAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAA 3452
OY      90      -----LysGlyGln--- 92
          -----
Db      3453      AAGAAAAAGAAAAAAATCATAGAAAAAGAAAAAGAAAAAGAAAAAGAAAGAGCCAGTGG 3512
OY      92      ----- 92
          -----
Db      3513      CAGTGAGCCAGGACAGTGAAGTGCCTCTCCCACTAGAAAGAAAGATAGTTAATTAGT 3572
OY      93      -----AsnGlyAsp-Asp----- 96
          -----|||::: |||
Db      3573      AATTTGTTATAAAAAGGAGAGAGACTAATATGAGACCACTTGAGACAGGTATGAGATGCC 3632
OY      97      -SerSerAlaGlyAlaPheProProProAlaGluValGluProThr---ProGluAl 115
          ::| ||||| ||| ||||| ||| ||| ||| |||
Db      3633      TGCATGGGCTGGAGGAAGCACTCCCTCCCGAGGCGCATGTGATCTGTGAGCCCACTCTTC 3692
OY      115      aGluLeu-----LeuLagInProCysSHSAspSe 125
          ||| |||
Db      3693      CTCCTCTCAGACTTCGCGTGTCTTGGAGGCCCCAGATAGCCATGAGCTCTGACTT 3752
OY      125      rGluAlaSerLysLeu-----GlyAl 132
          |||::: |||
Db      3753      CGGTGGTTCCAGACTTTCGCCATCAGACAGAGATCCATCAGAGATCAAGGACAGT 3812
OY      132      aProAlaAlaGlyGly-----GluGluGluIuTPG 142
          ||||| |||
Db      3813      CCGCGAGTGGCGATGACCTCAAGTTGTAACTATTAGGGGTTCTCGAAACCAAGAGCCAGGGG 3872
OY      142      Y---GInGlnGlnArgGluLeuGlyLysLysLysHisArgArgArgProSerLysLys 161
          ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      3873      CTGTTCGGGTGAGGGCGGTGCTGGCGCGAAGAACACCGTTAGGGCGCCATCGAAGCCCA 3932
OY      161      SAArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysLysPheAspGlu 181
          3933      GCGGACACTGGCGGCGCTACTTGGAGCTGAGCTGGGCGCGAAGAACACGACGATGTGAGAG 3992

```



```

OY 181 sGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProValAl 201
DB 3993 GCAGAGCAGGAGGCGCTCTCGGCTCGTAGAGATGTCGCAAAAGGTCACCCCTGGC 4052
OY 201 aProTyrAsnThrThrGlnPheMetAspAspHisAspGlnGluGlnProAspLeu 221
DB 4053 GCCCTACACACACACCGCTCTCATGATGATGCTGACCTGAGGAGGACCTTACCTC-- 4110
OY 221 sThrGlyLeuTyrSerLysArgAlaAlaLysSerAspAspThrSerAspAsp 241
DB 4111 -----GATGCGCTTCACGGGCGCTCCCA 4133
OY 241 eMetGluGluGlyGlyGlu--GluAspGlyLysSerAspGlyMetGlyLysAspGly 260
DB 4134 CTCTGCTCGGCTGGGAGAACGAGAGCGAGGAGGATGATGGGAGCCCGAGCCCATG 4193
OY 260 rGluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGlnSerLeu 280
DB 4194 GGAATTTACAGACGAGGAGCTTCTGAGGCGCTATGAGGAGTACACACTGAGAGCGCTTCA 4253
OY 280 nAsnMetSerLysGlnGluLeuLysGluTyrLeuGluGluLysCysLeuSer 300
DB 4254 GGGCCGAGCAGACGAGGAGCTGTCGAGACTGATCTAGAGAGCGCTGCTAC 4313
OY 300 gMetGluAspGluAsnAsnArgLeuArgLeuSerLysArgLeuGlyLysAspAsp 320
DB 4314 GGCTGAGCAGAACTCGGAGGCTCGG-----GAGCTGCGAGGTTGCTCCAG 4361
OY 320 aArg-----ValArgGluLeuGluLeuLeuAspArgLeuArgAla 335
DB 4362 CAGGACACCTGTCAACAGGTGAGAGACTGCTGCTGATGAGGAGGCTCAGACTGA 4421
OY 335 uAsnLeuGlnLeuLeuThrGluAsnGluLeuHisArgGln 349
DB 4422 AAACGAGCGGCTGCTCAGAGAACGAGATGTGGAACGAGAG 4464

RESULT 14
BC025970 1238 bp mRNA linear PRI 22-MAR-2002
DEFINITION Homo sapiens, similar to putative, clone MGC:39389 IMAGE:4559410,
ACCESSION BC025970
VERSION BC025970.1 GI:19684066
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1238)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (18-MAR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

NIH-MGC Project URL: http://mgc.ncl.nih.gov
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@cgsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Letitia Hsiao, Martin Kirzysinski, Keta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabh, Parvaneh Saeedi, Jacqueline
Schell, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 42 Row: d Column: 4  
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

FEATURES  
 source Location/Qualifiers  
 1..1238  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /db\_xref="taxon:9606"  
 /clone="MGC:39389 IMAGE:4559410"  
 /tissue="type="Eye, retinoblastoma"  
 /clone\_lib="NIH MGC\_16"  
 /lab\_host="DH10B-R"  
 /note="vector: pOTB7"  
 90..950  
 /codon\_start=1  
 /product="similar to putative"  
 /protein\_id="AAH25970.1"  
 /db\_xref="GI:19684067"  
 /translation="MATPNOTACAEAPVALBEAKTSGACGSPOTPPERRDSCGSLP  
 LTPRHSEDEBDLAGAVGLGWNRSRSTPGSCSAEVALARKHRRSRKRHW  
 RPYLELSVAEQOARDERQSORASRYREEMFAKGPVAPVNTTOFLMNDREDEPNLDV  
 PHGISHPGSGSESEAGSDSDGRRAHGEFRDFSEYRFTFSTESLQGRSKOELVLYL  
 ELERLSQAEETRLQLOACTGQSCQRYVELIAEYQRLRTENQRLROENQWNR  
 GCRDEEGRGT"

BASE COUNT 324 a 338 c 392 g 104 t

ORIGIN

Alignment Scores:  
 Pred. No.: 6.47e-21 Length: 1238  
 Score: 545.50 Matches: 131  
 Percent Similarity: 57.28% Conservative: 42  
 Best Local Similarity: 43.38% Mismatches: 99  
 Query Match: 28.56% Indels: 32  
 Gaps: 6

US-09-972-758a-2 (1-359) x BC025970 (1-1238)

```

OY 58 ArgProGlyProGluGlyGlySerLeuSerGlnProProProLeuGlnThrGln 77
DB 74 AGATTGGAGACAGAGATGATGCCA-CTCCGACACAGACCGCTGTAAATCAGAGTCA 132
OY 78 AlaCysProGluSerSerCysLeuArgGluGlyGln-----AsnGly 94
DB 133 CAGTGGCCCTG-GAGAGAGCCCAAGACCTCTGTGCCCCGGGAGCCCCCAACACCCCT 191
OY 95 AspAspSerSerAlaGlyLysAspPheProProAlaGluValGluProThrProGlu 114
DB 192 GAGCCTCATGACTCGTGGTGTCCCTGCTGACACCGCGAGGAGGACCTCAGAG 251
OY 115 AlaGluLeuLeuAlaGlnProCys-----HisAspSerGluAlaSerLysLeu 130
DB 252 GATGAAGATCTTGTGGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 311
OY 131 GlyAlaProAlaAlaGlyGlyGluGluGluTrpGlyGlnGlnGlnArgGlnLeuGlyLys 150
DB 312 CAGAGCCCGAGGGGCTGCTACGGGAG-----GCTGCTGCGCCCGG 353
OY 151 LysLysHisArgArgArgProSerLysLysArgHisTrpLysProTyrTyrLysLeu 170
DB 354 AAGAAACACCGCTCGCGCGCATCGAAGGCGCAAAAGCGACCTGACCCCTGAGACTG 413
OY 171 ThrTrpGluGluLysLysLysPheAspGluLysGlnSerLeuArgAlaSerArgIleArg 190
DB 414 AGCTGGGCTGGAACAAACACGCGGATGAGGAGGAGCAGAGGCGCTCCGCGTCCG 473
OY 191 AlaGluMetPheAlaLysGlyGlnProValAlaProTyrAsnThrThrGlnPheLeuMet 210
DB 474 GAAAGATGTTGGCAAAAGGCGACCCGCTGAGCCCTTACACACACCGCTCTCATG 533
OY 211 AspAspHisAspGlnGluPro-----AspLeuThrGlyLeuTyrSerLysArg 228

```







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 10:09:10 : Search time 72 Seconds  
(without alignments)  
664.402 Million cell updates/sec

Title: US-09-972-758A-2

Perfect score: 1910  
Sequence: 1 MAEPFLSEYOHOPOTSNCCTG.....LTENELHROERAPLSKFGD 359

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

A.Geneseq.101002: \*  
1: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1980.DAT: \*  
2: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1981.DAT: \*  
3: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1982.DAT: \*  
4: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1983.DAT: \*  
5: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1984.DAT: \*  
6: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1985.DAT: \*  
7: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1986.DAT: \*  
8: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1987.DAT: \*  
9: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1988.DAT: \*  
10: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1989.DAT: \*  
11: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1990.DAT: \*  
12: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1991.DAT: \*  
13: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1992.DAT: \*  
14: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1993.DAT: \*  
15: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1994.DAT: \*  
16: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1995.DAT: \*  
17: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1996.DAT: \*  
18: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1997.DAT: \*  
19: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1998.DAT: \*  
20: /SID22/gcgcdata/geneseq/geneseq-emb1/AA1999.DAT: \*  
21: /SID22/gcgcdata/geneseq/geneseq-emb1/AA2000.DAT: \*  
22: /SID22/gcgcdata/geneseq/geneseq-emb1/AA2001.DAT: \*  
23: /SID22/gcgcdata/geneseq/geneseq-emb1/AA2002.DAT: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1910	100.0	359	19	AAW85455
2	1910	100.0	359	23	ABP61799
3	1910	100.0	359	23	ABB76495
4	1903	99.6	359	22	AAW85465
5	308.5	16.2	152	22	AAW25726
6	237	12.4	134	22	ABG15306
7	209	10.9	349	22	ABB59598
8	155.5	8.1	281	22	ABG15303
9	147	7.7	954	22	AAU14615
10	143	7.5	288	22	ABG20362

11	141.5	7.4	1192	22	ABB59642
12	138	7.2	557	19	AAV20666
13	135.5	7.1	530	21	AAV94461
14	135.5	7.1	530	23	AAU74503
15	135.5	7.1	1749	22	ABG00839
16	135	7.1	554	23	ABB04721
17	135	7.1	555	22	AAW93869
18	134.5	7.0	510	22	ABB11764
19	134.5	7.0	510	22	AAW79741
20	134.5	7.0	546	22	AAW82808
21	133	7.0	538	21	AAW10043
22	133	7.0	648	20	AAV17946
23	133	7.0	1737	21	AAW10044
24	132.5	6.9	530	22	AAW64026
25	132.5	6.9	530	20	AAW97775
26	132.5	6.9	530	23	AAU74501
27	132.5	6.9	530	23	AAU74502
28	132	6.9	528	22	AAW78918
29	131	6.9	932	22	ABB65256
30	131	6.9	1424	22	AAW39253
31	131	6.9	1464	22	AAW41039
32	131	6.9	1898	20	AAV30795
33	130.5	6.8	819	22	AAW94316
34	130	6.8	611	20	AAV29039
35	130	6.8	611	22	AAU25510
36	129.5	6.8	1743	22	ABG10928
37	127.5	6.7	984	22	AAO12986
38	127	6.6	654	22	ABB63266
39	126.5	6.6	661	22	ABB61881
40	126.5	6.6	710	22	ABG20363
41	126.5	6.6	722	23	AAW47608
42	125.5	6.6	326	20	AAV20109
43	125.5	6.6	347	20	AAV31937
44	125.5	6.6	817	22	AAU31937
45	125	6.5	383	21	AAW13981

ALIGNMENTS

RESULT 1	AAW85455	standard; Protein; 359 AA.
ID	AAW85455	
XX	AAW85455	
AC	AAW85455	
XX	25-FEB-1999	(first entry)
DE	Secreted protein encoded by clone bp783_3.	
XX		
KW	Secreted protein; nutritional activity; immune stimulating; vaccine;	
KW	suppressing activity; haematopoiesis regulating activity;	
KW	tissue growth activity; activator; inhibitor activity; chemotactaxis;	
KW	chemokinetic activity; haemostasis; thrombolytic activity; receptor;	
KW	ligand; anti-inflammatory; cadherin; tumour invasion suppressor;	
KW	tumour inhibition; gene therapy.	
XX		
OS	Homo sapiens.	
XX		
PN	W09842739-A2.	
XX		
PD	01-OCT-1998.	
XX		
PF	20-MAR-1998:	98WO-US05653.
XX		
PR	19-MAR-1998:	98US-0044466.
XX		
PR	21-MAR-1997:	97US-0822167.
XX		
PA	(GENE ) GENETICS INST INC.	
XX		
PI	Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;	
XX	Racie LA, Spaulding V, Treacy M;	
XX		

Drosophila melanog  
Human neurofilamen  
Human p75 protein.  
Human nuclear tran  
Novel human diapo  
Human pp2464 prote  
Human polypeptide,  
Human LDL binding  
Human protein SFG  
Human low density  
MLV gag protein.  
MMLV gag gene pro  
MMLV gag-pol prote  
Human polypeptide  
Human polypeptide  
Human lens epithel  
Human nuclear anti  
Human lens epithel  
Human protein SFG  
Drosophila melanog  
Human polypeptide  
A human trichohyal  
Human protein sequ  
T. gondii immunoge  
Novel human diapo  
Drosophila melanog  
Drosophila melanog  
Novel human diapo  
Drosophila cell cy  
B. burgdorferi ant  
Novel human secret  
Arabidopsis thalia



```

Db      1 MAEPFLSEYOHQPOTSNCTGAAYQVEELNPERPPGAEERYPEEDSRMOSRAFPOLGGRPG 60
QY      61 PEGGSLSESOPPPLOTACPESSCLREGKQNGDDSSAGDPPPAVEPTPEAEILAAQ 120
Db      61 PEGGSLSESOPPPLOTACPESSCLREGKQNGDDSSAGDPPPAVEPTPEAEILAAQ 120
QY      121 PCHDSEASKIGAPAGEEEMGQOOLGKKHRRRPSKKRHMKPYKLTWEKKKFE 180
Db      121 PCHDSEASKIGAPAGEEEMGQOOLGKKHRRRPSKKRHMKPYKLTWEKKKFE 180
QY      181 KQSLRASIRIRAEKMPAKGQPYAPYNTTQFLMDHDHQPEDLKTGLYSKRAAKSDTSDDD 240
Db      181 KQSLRASIRIRAEKMPAKGQPYAPYNTTQFLMDHDHQPEDLKTGLYSKRAAKSDTSDDD 240
QY      241 FMEGGEEDGSDGMDGSEFLORDPSETYERYHTESLONMSKOELIKEYLEKCLSR 300
Db      241 FMEGGEEDGSDGMDGSEFLORDPSETYERYHTESLONMSKOELIKEYLEKCLSR 300
QY      301 MEDENNRLRLESKRLGDDARVRELLELDRLRAENQLLTENFLHQOERAPLSKRGD 359
Db      301 MEDENNRLRLESKRLGDDARVRELLELDRLRAENQLLTENFLHQOERAPLSKRGD 359

```

## RESULT 3

ABB76495 ID ABB76495 standard; Protein: 359 AA.

XX ABB76495:

DT 10-SEP-2002 (first entry)

DE Human oestrogen downregulated gene EDG1 protein.

XX EDG1: oestrogen downregulated gene; tumour suppressor; human;

KW breast cancer; prostate cancer; testicular cancer; ovarian cancer;

XX uterine cancer; colon cancer; gene therapy.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Region 150..177 /note= "nuclear localisation signal"

FT Misc-difference 154 /note= "encoded by AAG"

FT Misc-difference 171 /note= "encoded by AAC"

XX W0200228879-A1.

PD 11-APR-2002.

PF 05-OCT-2001; 2001WO-US31300.

PR 05-OCT-2000; 2000US-238187P.

XX (UYCA-) UNIV CASE WESTERN RESERVE.

XX Montano M, Wittman B;

XX WPI: 2002-519107/55.

XX N-PSDB; ABN84013.

XX Polynucleotides encoding Estrogen Down-Regulated Gene 1 proteins,

XX useful for the prevention, diagnosis and treatment of e.g. breast

XX cancer, testicular cancer, prostate cancer, uterine cancer, cervical

XX cancer, ovarian cancer and colon cancer -

XX Claim 7; Fig 1A-B; 52pp: English.

XX The present sequence is the protein sequence of human EDG1, a

XX 40 kDa protein encoded by the tumour suppressor gene designated

XX CC human oestrogen downregulated gene 1 (see ABN84013) EDG1 mRNA

XX expression is prevalent in normal mammary epithelial cells and in

```

CC other human hormone-responsive tissues such as the ovary, prostate
CC and testis. Expression is low in breast cancer epithelial cells.
CC Oestradiol, which induces breast cancer cell growth, has an
CC inhibitory effect on EDG1 mRNA expression in breast cancer cells.
CC Hexamethylene bis-acetamide, an inducer of differentiation and
CC apoptosis, upregulates EDG1 mRNA expression in breast cancer cells.
CC The invention provides EDG1 polynucleotides and polypeptides. In
CC a claimed method of detecting cancerous cells, a test sample from
CC an individual suspected of having, or known to have breast,
CC testicular, prostate, uterine, cervical, ovarian or colon cancer is
CC contacted with an anti-EDG1 antibody. A decrease in the level
CC or antigen-antibody complex compared to the level of a control
CC sample indicates cancerous cells. A claimed method for decreasing
CC the proliferation of breast, prostate, testicular, ovarian, uterine,
CC cervical or colon cancer cells involves increasing EDG1 protein
CC activity in the cells, either by contacting the cells with EDG1
CC protein, its fragment or functional equivalent, or with a nucleic
CC acid encoding EDG1 protein, its fragment or functional equivalent.
XX
SQ Sequence 359 AA:

```

```

Query Match 100.0%; Score 1910; DB 23; Length 359;
Best Local Similarity 100.0%; Pred. No. 5.9e-153;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 MAEPFLSEYOHQPOTSNCTGAAYQVEELNPERPPGAEERYPEEDSRMOSRAFPOLGGRPG 60
Db      1 MAEPFLSEYOHQPOTSNCTGAAYQVEELNPERPPGAEERYPEEDSRMOSRAFPOLGGRPG 60
QY      61 PEGGSLSESOPPPLOTACPESSCLREGKQNGDDSSAGDPPPAVEPTPEAEILAAQ 120
Db      61 PEGGSLSESOPPPLOTACPESSCLREGKQNGDDSSAGDPPPAVEPTPEAEILAAQ 120
QY      121 PCHDSEASKIGAPAGEEEMGQOOLGKKHRRRPSKKRHMKPYKLTWEKKKFE 180
Db      121 PCHDSEASKIGAPAGEEEMGQOOLGKKHRRRPSKKRHMKPYKLTWEKKKFE 180
QY      181 KQSLRASIRIRAEKMPAKGQPYAPYNTTQFLMDHDHQPEDLKTGLYSKRAAKSDTSDDD 240
Db      181 KQSLRASIRIRAEKMPAKGQPYAPYNTTQFLMDHDHQPEDLKTGLYSKRAAKSDTSDDD 240
QY      241 FMEGGEEDGSDGMDGSEFLORDPSETYERYHTESLONMSKOELIKEYLEKCLSR 300
Db      241 FMEGGEEDGSDGMDGSEFLORDPSETYERYHTESLONMSKOELIKEYLEKCLSR 300
QY      301 MEDENNRLRLESKRLGDDARVRELLELDRLRAENQLLTENFLHQOERAPLSKRGD 359
Db      301 MEDENNRLRLESKRLGDDARVRELLELDRLRAENQLLTENFLHQOERAPLSKRGD 359

```

## RESULT 4

AAB95465 ID AAB95465 standard; Protein: 359 AA.

XX AAB95465:

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:17953.

XX Human; primer: detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EPI074617-A2.

XX 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI: 2001-318749/34.  
 XX  
 PT primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 8; SEQ ID 17953; 2537pp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination  
 CC of the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
 CC AAH95693 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 CC  
 CC  
 SO Sequence 359 AA:

Query Match 99.6%; Score 1903; DB 22; Length 359;  
 Best Local Similarity 99.7%; Pred. No. 2.3e-152;  
 Matches 358; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

YY 1 MAEPFLSEYOHOPOTSNCTGAAAOEELNPRPGAEPRPEEDSRMSRAFPOLGGPG 60  
 Db 1 MAEPFLSEYOHOPOTSNCTGAAAOEELNPRPGAEPRPEEDSRMSRAFPOLGGPG 60  
 YY 61 PEGGSLSEOPPLQOTGACPESSCLREGCKGQNGDSSAGGDFPPPEVEPTPEALLAQ 120  
 Db 61 PEGGSLSEOPPLQOTGACPESSCLREGCKGQNGDSSAGGDFPPPEVEPTPEALLAQ 120  
 YY 121 PCHSESESKGAPAGGEEWGOOQOLGKKHRRRPSKKRHKPKYKLTWEKKKFEDE 180  
 Db 121 PCHSESESKGAPAGGEEWGOOQOLGKKHRRRPSKKRHKPKYKLTWEKKKFEDE 180  
 YY 181 KOSIRASIRIREREMAKGOPYAPYNTTQFLMDHDOEEDLKTGLYSRAAKSDTSDDD 240  
 Db 181 KOSIRASIRIREREMAKGOPYAPYNTTQFLMDHDOEEDLKTGLYSRAAKSDTSDDD 240  
 YY 241 FMEGGEDEGSDGMDGSEFLORDFSETYERHTESLONMSKOEILKEYLEKLSR 300  
 Db 241 FMEGGEDEGSDGMDGSEFLORDFSETYERHTESLONMSKOEILKEYLEKLSR 300  
 YY 301 MEDENNRLRLSKRLGDDARVRELLELDRLAENQLLTENFLHQOERAPYKRTGD 359  
 Db 301 MEDENNRLRLSKRLGDDARVRELLELDRLAENQLLTENFLHQOERAPYKRTGD 359

ID AAM25726 standard; Protein: 152 AA.  
 XX  
 AC AAM25726:  
 XX  
 DT 16-OCT-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:1241.  
 XX  
 KW Human; cancer; HIV infection; human immunodeficiency virus;  
 KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
 KW antibacterial; endocrine; cardiac; central nervous system; virucide;  
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
 KW antiagregant; haemostatic; vulnery; antileuc; osteopathic; eczema;  
 KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;  
 KW neuroproliferative; antidepressant; nootropic; antiparkinsonian; infection;  
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
 KW antiaphyllactic; rheumatoid arthritis; septic shock; pancreatitis;  
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
 KW thrombocytopenia; osteoporosis; severe combined immunodeficiency;  
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
 KW neurological disorder.  
 KW  
 XX Homo sapiens.  
 OS  
 PN WO200153455-A2.  
 XX  
 PD 26-JUL-2001.  
 XX  
 PF 22-DEC-2000; 2000WO-US35017.  
 XX  
 PE 23-DEC-1999; 99US-0471275.  
 PR 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-0552317.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT;  
 XX  
 DR WPI: 2001-457603/49.  
 DR N-PSDB; AAH99667.  
 XX  
 PT Isolated human polynucleotides encoding polypeptides, useful for the  
 XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection -  
 XX  
 PS Claim 20; Page 257; 1217pp; English.  
 XX  
 CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to  
 CC AAM25963. The proteins can have activities based on the tissues and  
 CC cells they are expressed in, such as: antiinflammatory; antirheumatic;  
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;  
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;  
 CC cardiovascular; antianaemic; antiagregant; haemostatic; vulnery;  
 CC antidiuretic; osteopathic; dermatological; antiallergic; antiasthmatic;  
 CC antidiabetic; cytostatic; neuroproliferative; antidepressant; nootropic;  
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
 CC encoding them can be used in gene therapy, antisense therapy and vaccine  
 CC production. The proteins and polynucleotides are useful for screening for  
 CC agonists or antagonists of a protein and for the treatment and diagnosis  
 CC of disorders associated with the activity of a protein e.g. inflammation,  
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,  
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,  
 CC anaemia, platelet disorders, thrombocytopenia, wounds, burns, ulcers,  
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic  
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and  
 CC neurological disorders.  
 XX  
 SO Sequence 152 AA;

Query Match 16.2%; Score 308.5; DB 22; Length 152;



Query Match	Similarity	Score	DB	Length
Best Local	25.9%	Pred. No. 2,4e-09;		
Matches	68;	Conservative	43;	Mismatches 84; Indels 68; Gaps 10
OY	85	LREGEKGND-----DSAGGDFPPPAVEPTPEAEILLAPCHDSEASKLGPAPAGEE	139	
Db	1	MAEAVKNESGSOQRRLDSDGGGG-----ASGGGVAVAGGS	36	
OY	140	EMGQOQOROLGKKHRRRSKKRHHKPYKLTWE-----EKKFDEKQSLRARIAREMA	195	
Db	37	GMPRKKHRRGKS-KMOPKTKNHY-PQKKLMSGAGATYLGNOONSRKTLVRSRL-	93	
OY	196	KGQPAVPNTTQFLMDHDDEBDPLKTGLYSKRAAKSDPTDDDFMEEGGEDGSDGM	255	
Db	94	---LVPYNTNRFMEEHMSF-----LHRDSDDNCF---GQTE-----	126	
OY	256	GGDGSEFLQRPDSFERYRHTESLQMSKQELIKLEYLELEKLSRMEDNNRLRESKL	315	
Db	127	--DQVLFSTKESDYERARLERLETMSKQELIOECMOIEDRYSKAONTSKEF---GAKL	181	
OY	316	GGDARVRELELELDRLRAENIQ	338	
Db	182	RAQDDKIRQLSRENOFLRTHLLR	204	
RESULT 8				
ABG15303				
ID	ABG15303	standard; Protein: 281 AA.		
XX	ABG15303;			
AC				
XX	18-FEB-2002	(first entry)		
DT				
XX				
DE		Novel human diagnostic protein #15294.		
XX				
KW		Human; chromosome mapping; gene mapping; gene therapy; forensic;		
XX		food supplement; medical imaging; diagnostic; genetic disorder.		
OS		Homo sapiens.		
XX				
PN		MO200175067-A2.		
XX				
PD		11-OCT-2001.		
XX				
PF		30-MAR-2001; 2001WO-US08631.		
XX				
PR		31-MAR-2000; 2000US-0540217.		
XX				
PR		23-AUG-2000; 2000US-0649167.		
XX				
PA		(HYSE-) HYSEQ INC.		
XX				
PI		Drmanac RT, Liu C, Tang YT;		
XX				
DR		WPI: 2001-639362/73.		
XX				
DR		N-PSDB; AAS79490.		
XX				
PT		New isolated polynucleotide and encoded polypeptides, useful in		
XX		diagnostics, forensics, gene mapping, identification of mutations		
PT		responsible for genetic disorders or other traits and to assess		
XX		biodiversity.		
PS		Claim 20; SEQ ID NO 45662; 103bp; English.		
XX				
CC		The invention relates to isolated polynucleotide (I) and		
CC		polypeptide (II) sequences. (I) is useful as hybridisation probes,		
CC		polymerase chain reaction (PCR) primers, oligomers, and for chromosome		
CC		and gene mapping, and in recombinant production of (II). The		
CC		polynucleotides are also used in diagnostics as expressed sequence tags		
CC		for identifying expressed genes. (I) is useful in gene therapy techniques		
CC		to restore normal activity of (II) or to treat disease states involving		
CC		(II). (III) is useful for generating antibodies against it, detecting or		
CC		quantitating a polypeptide in tissue, as molecular weight markers and as		
CC		a food supplement. (II) and its binding partners are useful in medical		

[illegible]



XX Claim 10; Page 215-217; 392pp; English.

PS ANU14602-ANU14794 represent novel bone marrow polypeptides of the  
XX invention. The proteins and corresponding coding sequences may be used  
CC in the prevention, diagnosis and treatment of diseases associated with  
CC inappropriate bone marrow polypeptide expression. For example, to treat  
CC disorders associated with decreased expression by rectifying mutations  
CC or deletions in a patient's genome that affect the activity of the  
CC polypeptides by expressing inactive proteins or to supplement the  
CC patient's own production of the polypeptide. Additionally, the nucleic  
CC acids may be used to produce the polypeptides, by inserting the nucleic  
CC acids into a host cell and culturing the cell to express the protein.  
CC The nucleic acid and its complementary sequences may also be used as DNA  
CC probes in diagnostic assays to detect and quantitate the presence of  
CC similar nucleic acid sequences in samples, and therefore which patients  
CC may be in need of restorative therapy. The proteins may also be used as  
CC antigens in the production of antibodies against bone marrow proteins  
CC and in assays to identify modulators of their expression and activity.  
CC The anti-bone marrow protein antibodies and antagonists may also be used  
CC to down regulate expression and activity. The antibodies may also be used  
CC as diagnostic agents for detecting the presence of the protein in samples  
CC (e.g. by enzyme linked immunosorbent assay (ELISA)). The proteins  
CC may be used to regulate haematopoiesis activity, and consequently in the  
CC treatment of myeloid or lymph cell disorders; in tissue regeneration,  
CC such as wound healing; as a nutritional supplement; and in treatment of  
CC immune disorders such as severe combined immunodeficiency (SCID).

XX Sequence 954 AA:

Query Match 7.7%; Score 147; DB 22; Length 954;  
Best Local Similarity 21.6%; Pred. No. 0.0015;

Matches 97; Conservative 65; Mismatches 146; Indels 142; Gaps 19;

QY 25 QBELNPPPPAEEER-----VEPDSRWOS--RAF-----POLGRRPCEGE----- 64  
DB 189 QDELQSKOSKGLLEEYHRLSPPLPHEE--RAQSPRLATEPEPPQ-----GPEGPEWKE 242  
QY 65 -----GSLSEPPPIQTOA--CPRESSCLREGKQNGCNDSSAGCPPPPAE-----VE 110  
DB 243 AEELGDSASLSLQSLQREDAVSPAPACEKGEKHQSOAEELGGEAEADPEKAVVS 302  
QY 111 PTP-----BAELLAOPCHDSEASKLGAAPAGEEEMGDOOROLGKKK----- 152  
DB 303 PTPVSPPEVRSTEPYAPPEQLSEAA--LKAMEEAVNQVLEQDQRLHLESKQEMQGLRKL 361  
QY 153 -----HRRPSSKKRRHKMPYKLTWEKKKFFPEKQSLASRSIRAEKFAKGPYA 201  
DB 362 CQEEEEETLRLHQKEQSLREHLOKALIEEERARMEESORLSWLRAGVOSTQA-- 419  
QY 202 PYNTTQFLMDHDEEPLKTGLYSKRAAKSD-----DTSDDDMEGEGEDGGS- 252  
DB 420 -----DEDOIARAQDEASLQKLREELSQOKAEKVASLQFQKNRMQLDEKTEIASE 469  
QY 253 -----DGMGDGSSEFLQRFSEYERY-----HTE-----SLQ 280  
DB 470 KSEQALNAKEKALQQLREQLGERKPAVATLEKESHALEJRLCSSLEAKHREVVSLSQ 529  
QY 281 NMSQGLKEYLELEKCLSRMEDENNRLLESKRLGSGDARVREL-----ELE 328  
DB 530 KRIEQAQKEBAQLQKQGYVE--HRYHOKSYHVAQYEHLSLSLREKROEVEGEHRR 586  
QY 329 IDRLRAENLQILIT-----ENELHROOERAPL 354  
DB 587 LDKMKHEHQYMAKAREQYFAAEERQRAEL 616

RESULT 10  
ABG20362  
ID ABG20362 standard: Protein: 288 AA.

XX AC ABG20362;  
XX

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #20353.

KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PP 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB: AAS84549.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

PS Claim 20; SEQ ID No 50721; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 288 AA:

Query Match 7.5%; Score 143; DB 22; Length 288;  
Best Local Similarity 27.5%; Pred. No. 0.0007;

Matches 69; Conservative 24; Mismatches 64; Indels 94; Gaps 14;

QY 50 RAPQUL--GGRPP-----PEGESLSQPPPIQTOACPESSCLREGKQGNDDDS 97  
DB 9 RAGVOLNQGARGPGSTAGDACRRATGTL-----ALRLRA-----RSEKRODSALER 56  
QY 98 SAGGDFPPPAEVEPTPEAELLAPCHDSEASKLGAAPAGE-----EEM-GQOOR---- 146  
DB 57 S--GRAPPPL-----LRGLGANGAGARGGCGCDRWGCGQWRPGRF 95  
QY 147 -----QLGKKKRRRPSKKRRHKMPYKLTW-----BEKKKFPDK----- 181  
DB 96 EIKKKLTKAKKKKKKEEODEKKKLTQIOESOVTSNNKRRSKRDKLDKKSQAM 155  
QY 182 QSLRASRIR-----AEMFAKQGPVAPYNTTQFLMDHDEEPLDKTGLYSKRAAKSDDT 236

Seq ID	Seq Name	Seq Type	Seq Length	Seq Description
Db	156	EEIKERERKRRKRTTAEKAKKPL	---	KTSEVYSDDDEEEEDD-----KSKESDRSSRT 207
Oy	237	SDDDFMEEGE	247	
Db	208	SSSDEEEKKE	218	
RESULT 11				
ID	ABB59642	standard; Protein; 1192 AA.		
AC	ABB59642;			
DT	26-MAR-2002	(first entry)		
XX	Drosophila melanogaster polypeptide SEQ ID NO 5718.			
XX	Drosophila; developmental biology; cell signalling; insecticide;			
XX	pharmaceutical.			
XX	Drosophila melanogaster.			
XX	WO200171042-A2.			
XX	27-SEP-2001.			
PF	23-MAR-2001; 2001WO-US09231.			
PR	23-MAR-2000; 2000US-191637P.			
PR	11-JUL-2000; 2000US-0614150.			
XX	(PEKE ) PE CORP NY.			
PI	Venter JC, Adams M, Li PWD, Myers EW;			
DR	WPI: 2001-656686/75.			
DR	N-PSDB; ABL03745.			
PT	New isolated nucleic acid detection reagent for detecting 1000 or more			
PT	genes from Drosophila and for elucidating cell signalling and cell-cell			
PT	interactions -			
XX	Disclosure: SEQ ID NO 5718; 21pp + Sequence Listing; English.			
XX	The invention relates to an isolated nucleic acid detection reagent			
CC	capable of detecting 1000 or more genes from Drosophila. The invention is			
CC	useful in developmental biology and in elucidating cell signalling and			
CC	cell-cell interactions in higher eukaryotes for the development of			
CC	insecticides, therapeutics and pharmaceutical drugs. The invention			
CC	discloses genomic DNA sequences (ABLI6176-ABL30511), expressed DNA			
CC	sequences (ABLI57737-ABR2072).			
CC	The sequence data for this patent did not form part of the printed			
CC	specification, but was obtained in electronic format directly from WIPO			
CC	at ftp.wipo.int/pub/published_pct_sequences.			
XX				
XX				
SO	Sequence 1192 AA;			
Oy	Query Match	7.4%;	Score 141.5;	DB 22; Length 1192;
Db	Best Local Similarity	22.3%;	Pred. No. 0.006;	
Db	Matches	80; Conservative	57; Mismatches	138; Indels 83; Gaps 17
Oy	4	PFLSYGQOPQTSNCTGAANOELINPERPPAEERVPEDSRMOSRAFPOLGRRPG	63	
Db	252	PLNDELPEDESSPAATESAV--EELEKESAMMDQVEPESEIOEOVOP--GEYOSDS	306	
Oy	64	EG-----SESOPLPLOTACPPSSCLREGEKQNGDSSAGDFFPPAVEPTPEAE	116	
Db	307	DGEQAEKPELEAQF---EVEAQPAEAQPEAF-----PQLEVEFOPEVE	348	
Oy	117	LLAOPCHDSAEASKGAPAAAGGEEMGQOORLGKKKRRRRSPKSKKRWKPYVYKLTWEKK	176	

Db	34-9	-SQPEVESQPEVEVAQPEVEPOSEVESQ-----PEAESHSPE---TQAEVE	390
Qy	177	KFEKQSLRASHIRAEFMFAKQPVAPYNTTQFLMDHDHDEEPDLKTGLYSKRAAKSDT	236
Db	391	AQPEVESLPEAESQPE--AESQP-----EREPEVE---AEKISDNEVD	429
Qy	237	SDDPMEEGCGEFGDGGSDGNGGCGSEFLORDPSETERYRTTESLQNNKSKELIKYELEK	296
Db	430	TEASIMETLIVE--GIEDGILTAAMDNIYPEELAEASDKQETE--LESDDQSPYTEALE-EQ	485
Qy	297	CLSRMEDNNRLRLSKRLGCDGARVRELELELDRLRAENLC-LTENEELHQOEAP	353
Db	486	AVPEIEQEKER---EPEQITLAD-----ETQDSQAQPSNEEPVELAPQHTAEALAP	534
RESULT 12			
AAV20666			
ID	AAV20666	standard; Protein: 557 AA.	
XX	AAV20666;		
XX			
DT	22-JUL-1999	(first entry)	
XX			
DE	Human neurofilament-M wild type protein fragment 8.		
XX			
KM	Human; beta-amyloid precursor protein; beta-ApP; diagnosis: cancer;		
KM	frameshift mutation; age-related disease; neurodegenerative disorder;		
KM	Alzheimer's disease; Down's syndrome; myotonic dystrophy; neuronal;		
KM	Huntington's disease; multiple sclerosis; alcoholic liver disease;		
KM	diabetes mellitus type II; microtubule associated protein; Tau; Big Tau;		
KM	ubiquitin B; apolipoprotein E; MAP2; neurofilament-L; neurofilament-M;		
KM	neurofilament-F; presenilin 1; presenilin II; cellular tumour antigen;		
KM	glial fibrillary acidic protein; GFAP; p53; semaphorin III; HUPF-1;		
KM	bcl-2; B-cell leukemia/lymphoma 2 proto-oncogene; HMGP-C; NSP-A;		
XX	high mobility group protein-C; neuroendocrine specific protein A.		
XX			
OS	Homo sapiens.		
XX			
PN	W09845322-A2.		
XX			
PD	15-OCT-1998.		
XX			
PF	02-APR-1998; 98WO-IB00705.		
XX			
PR	10-APR-1997; 97US-0043163.		
XX			
PA	(UYUT-) RIJXSUNIV UTRECHT.		
PA	(ROTA-) ROTAL NETHERLANDS ACAD ARTS & SCI.		
PA	(UYRO-) UNIV ROTTERDAM ERASMUS.		
XX			
PI	Burdach JPH, Grosveid FG, Van Leeuwen FW;		
XX			
DR	WPI. 1998-609901/51.		
XX	N-PSDB: AAX75759.		
XX			
PT	Diagnosing disease by detecting frameshift mutations in RNA or		
PT	corresponding protein mutations - used to diagnose cancer and		
PT	neurological diseases, particularly Alzheimer's disease, and also		
PT	for treatment and prevention with specific ribozymes or wild-type		
PT	RNA		
XX			
PS	Disclosure; Figure 8; 258pp; English.		
XX			
CC	This invention describes a novel method for the diagnosis of a disease		
CC	caused by, or associated with, an RNA molecule that has a frameshift		
CC	mutation. The method is used to diagnose age-related diseases, especially		
CC	cancer and a wide range of neurodegenerative disorders (e.g. Alzheimer's		
CC	disease, Down's syndrome, myotonic dystrophy, Huntington's disease,		
CC	multiple sclerosis, alcoholic liver disease, diabetes mellitus type II		
CC	and many others listed) or susceptibility to these disorders. The method		
CC	allows a definitive diagnosis of Alzheimer's disease in living patients,		
CC	at an early stage. It is based on the observation that disease may be		
CC	caused by mutations in RNA rather than DNA. The invention describes the		







**THIS PAGE BLANK (USPTO)**